Publishing Authoritative Irish Geospatial Data to Support Interlinking of Building Information Models

Editor(s): Name Surname, University, Country
Solicited review(s): Name Surname, University, Country
Open review(s): Name Surname, University, Country

Kris McGlinn[a], Rob Brennan[a], Christophe Debruyne[a], Alan Meehan[a], Lorraine McNerney[b,], Eamonn Clinton[b], Philip Kelly[b] and Declan O'Sullivan[a]
[a] *ADAPT, Trinity College Dublin, Dublin 2, Ireland*
[b] *Ordnance Survey Ireland, Dublin 8, Ireland*

Abstract. Building Information Modelling (BIM) is a key enabler to support integration of building data within the buildings life cycle (BLC) and is an important aspect to support a wide range of use cases, related to intelligent automation, navigation, energy efficiency, sustainability and so forth. Open building data faces several challenges related to standardization, data interdependency, data access, and security. In addition to these technical challenges, there remains the barrier among BIM developers who wish to protect their intellectual property, as full 3D BIM development requires expertise and effort. This means that there is often limited availability of building data. However, a Linked Data approach to BIM, combined with a supporting national geospatial identifier infrastructure makes interlinking and controlled sharing of BIM models possible. In Ireland, the Ordnance Survey Ireland (OSi) maintains a substantial data set, called Prime2, which includes not only building GIS data (polygon footprint, geodetic coordinate), but also additional building specific data (e.g. form, function and status). The data set also includes change information, recording when changes took place and who captured and validated those changes. This paper presents the development of a national geospatial identifier infrastructure based on an OSi building ontology that supports capturing OSi building data as RDF. The paper details the different steps required to generate the ontology and publish the data. First, an initial analysis of the data set to generate the ontology is discussed. This includes identification of mappings to existing standards, e.g. GeoSPARQL to handle geometries and PROV-O to handle provenance, to the development of R2RML mappings to generate the RDF and the method for deploying the ontology and the building graphs. This data is then made available dependent on different licensing agreements handled by an access control approach. Methods are then presented to support the interlinking of the authoritative data with other building data standards and data sets using geolocation, followed finally by discussion and future work.

Keywords: Building Information Modelling, Geographic Information Systems, Ontology Engineering, Resource Description Framework (RDF); Linked Data;

1. Introduction

Access to reliable structured data plays a central role in supporting existing and future services for managing smart and sustainable buildings and cities [1]. ICT solutions are becoming increasingly important for data integration [2] and for supporting new control and monitoring capabilities for managing buildings. Building Information Modelling (BIM) has been identified as a key enabler to support integration of building data not only within the buildings life cycle (BLC), which includes its design, construction, operation and re-design (e.g. renovation), as well as demolition/recycling [3] [4], but also with other data sources, such as those related to geolocation, people and their behavior, weather, energy, etc. [5]. Open building data has the potential to support new and innovative services to support intelligent automation, navigation, energy efficiency and sustainability, but faces several challenges, related to standardization, data interdependency, data access, and security [6]. In addition to these technical challenges, there remains the barrier amongst building data owners who wish to protect their intellectual property, as full 3D building model development requires expertise and effort [7]. This means that there is often limited availability of data about buildings. One way to bridge the gap between individual building models and the broader spatial environment, and to enable building data interlinking, is to create a common digital spatial infrastructure upon which building data can be consistently integrated, referenced and perhaps retrieved [8], [9]. Geospatial Linked Data is a technical mechanism to achieve this goal. Publishing data is not enough. The fundamental service of such a system is the sensing, classification, verification and maintenance of a consistent building identifier system over the long term.

In 2014, Ordnance Survey Ireland (OSi, Ireland's national mapping agency) delivered a newly developed spatial data storage model known as Prime2 [11]. With Prime2, OSi moved from a traditional map-centric model to an object-oriented model from which various types of mapping and data services are produced. Prime2 currently holds information of over 50 million uniquely identified spatial objects (road segments, fences, rivers, lakes etc.); of which some have more than one geometric representation. This includes over 3.5 million buildings. Each building has GIS data (polygon footprint and geodetic coordinate) and additional data such as form, function and status, as well as provenance data related to changes made to the building over time. Each Prime2 spatial object had a unique 128-bit identifier assigned (GUID/UUID) for internal tracking. A challenge was converting these to easily accessed and maintained external identifiers.

This paper presents the development of the OSi building ontology and the generation of the OSi building data as Linked Data (LD) [12]. As this LD is authoritative, it provides an excellent basis for interlinking with other building and building related data sets supporting the iterative development of ever more complex and integrated building models. The paper is structured as follows. Section 2 presents the background and related work, with a focus on the important role of standardization to support interoperability. Section 3 describes the methodology for developing the ontology and generating the RDF data. For each step in the methodology, how it is applied to the OSi building data schema is described. This consists of an initial analysis of the data which includes identification of mappings to existing standards, e.g. for the OSi building data, GeoSPARQL [13] to handle geometries and PROV-O [14] to handle provenance. In section 5 the ontology development is described. This is used as a basis for the definition of R2RML mappings to generate the RDF using open software, which is then presented in section 6.

Section 7 presents a method for allowing access rights dependent on different licensing agreements is then presented, to support selective access to data, an important requirement for building data which may be sensitive, or have value which can be monetized. This is followed by section 8, a presentation of use cases to demonstrate the use of the published data to support the interlinking of the authoritative data with other BIM standards (Industry Foundation Classes [15]) and data sets (DBpedia [16], open governmental data and Irish Central Statistics Office data [17]) based around the authoritative URIs and geolocation, thus providing a basis for interlinking and integrating open building data. Finally, the conclusion and future work is presented.

2. Background and Related Work

As the Prime2 data set is a geospatial data set, we first explore Geographical Information Systems (GIS) and the CityGML standard for representation these types of data. Next Building Information Modelling (BIM) is discussed, and the Industry Foundation Classes (IFC) standard. Finally, approaches to the use of Linked Data (LD) and ontologies to support structured data, accessible over the web are discussed.

2.1. Geographical Information Systems

GIS systems are information systems with an added geo-reference [18][19]. The geospatial information associated with GIS systems allow spatial analysis to be performed. GIS data has a vast array of use cases including the monitoring of weather systems across a region, predicting population densities in a city and visualising real time traffic jams in a certain location. Analysis of GIS data gives important location-based insights which may have previously been overlooked. Geospatial information in a GIS system typically includes the coordinates of features of an object, based on the real-world location (geolocation) of the object. The relationship between features of the object can also be defined. Such features may include the walls of a building, and how certain walls relate to one another by being associated with a room within the building. GIS features are commonly represented in a

2D coordinate system. A building would be represented by its top-down building footprint. A GIS system representation of a building will also include the building within the context of other geospatial features, such as infrastructure and natural features. This makes the information relevant to building designers and public planning organisations, planning for navigation, fire services, energy grids, etc.

2.1.1. CityGML: A Standard for GIS

Recent research in GIS primarily focuses on the development of 3D geospatial models [18]. Such systems facilitate the modelling of complex internal structures of buildings, tunnels and bridges in the geospatial domain. City Geography Markup Language (CityGML) is currently the state of the art standard in 3D GIS modelling, providing an XML based data model for the storage and exchange of 3D models [20]. CityGML is implemented as an application schema for the Graphical Modelling Language (GML), by the Open Geospatial Consortium (OGC), and it allows capturing 3D models of cities and landscapes. This includes the description of objects geometry, topology, semantics and appearance between thematic classes, aggregations, relations between objects, and spatial properties. This 3D GIS modelling approach is closer to Building Information Modelling (BIM) than 2D GIS, however work still remains in providing seamless integration of these two representations [8]. In CityGML, buildings are described at five levels of detail (LOD), and at its most detailed level, this includes descriptions of rooms, furniture, openings and installations (lamps, radiators). It usually does not have the kind of detail found in an Industry Foundation Classes (IFC) data model though (see 2.2.1).

2.2. Building Information Modelling

The concept of Building Information Modelling (BIM) has been created to support the vast amount of data associated with buildings. This data is generated across the building's life cycle (BLC) and requires maintenance and management throughout the BLC. BIM describes an integrated data model for storing all information relevant to the BLC, typically relating to the functional and physical characteristics of a building [4]. This primarily includes a 3D model of the architectural design, detailing the positions and dimensions of a building's walls, rooms, windows, doors, roof, etc. BIM also facilitates the inclusion of non-physical building features such as the building costs, accessibility, safety, security and sustainability [21]. BIM is therefore capable of capturing all aspects of a building that exist throughout the BLC, aiding stakeholders at all stages of the BLC. As the authors state in [18], "it is clear that BIM is not just a piece of software, but also a process that contributes to the workflow and project delivery process."

The use of BIM is active and growing in Ireland, with a 2016 Irish Digital Transition Survey reporting that 76% of respondents possess confidence in their organization's BIM skills and knowledge [22]. Ireland is looking to follow the UK process, which has had a strong drive to generate Level 2 BIM for all centrally procured projects in England, Wales and Northern Ireland. Like the UK, challenges remain in Ireland for SMEs who must weigh the known benefits against barriers, such as costs of software and training [23]. The availability of open building data is also subject to these same issues in Ireland, as across the globe, i.e. that there are still many barriers to sharing building data, related to standardization, data interdependency, data access and security [6].

To better support existing and future use cases, it is important that developers be given access to available, open and authoritative (i.e. trustworthy) building data. Currently, GIS and other data sources, can be used to construct rudimentary building data models, based on location, and other attributes like address, are available openly. In Ireland, these data are scattered between different services including the aforementioned data.gov, as well as DBpedia. Linking these data sources with an authoritative building data dataset, provides an important step toward making building data available and the use of open standards is a necessary requirement to support this process. Standardization of datasets is an important part of ensuring data interoperability [24].

2.2.1. Industry Foundation Classes

Within the Architecture, Engineering and Construction (AEC) community, the leading standard around the concept of BIM is Industry Foundation Classes (IFC), developed by buildingSmart [25]. IFC is also the only standards for exchanging building information which is also an ISO PAS standard [15], and so it remains a primary candidate for exchanging building data. IFC is a non-proprietary data model, which addresses several core data domains, required for building AEC processes (architecture, structural analysis, control, etc.), enabling information to be passed between different stakeholders across the BLC. IFC has seen major government clients in the UK, Norway, and Finland, as well as a growing commitment in China [26] and the US [27]. IFC is based upon the EXPRESS schema and maintains a complex set of relationships. This complexity can be a barrier to non-experts, for example, web developers, who may want to make use of available models to support their applications, be they related to navigation, building controls, sustainability, etc. This combined with a reluctance of owners of IFC models to share those models, due to for example security concerns or protection of intellectual property (developing a full IFC model can be a labour-intensive task), means that the current situation is a severe lack of openly available IFC models available to developers.

2.3. Linked Data to Support Interlinking

Linked Data (LD) is an approach to expose, share, and connect related data, which was not previously linked, on the Web [12]. RDF and textual (HTML) content do not just live next to each other on the Web of Data but are also indirectly connected to each other. In modern AEC, data related to different domains such as building geometry and topology data, sensor data, behaviour data, geo data, are generated and consumed across BLC stages. The representation of building data as linked data has the potential to meet the requirements for storing and sharing those data. However, those data have to be represented as or at least tagged using RDF. With the development of the ifcOWL [28] standard in buildingSMART [29], this is now possible.

IfcOWL transforms the well-established IFC standard, defined in EXPRESS schema, into OWL opening up the potential for linking ifcOWL, and other ontologies for representing building data, with other domains such as Smart Appliances Reference ontology (SAREF) [30], DogOnt [31], the Semantic Sensor Network (SSN) and SOSA (Sensor, Observation, Sample, and Actuator) for sensor devices domain[1], etc.. An approach has also been developed to transform GbXML into OWL [32] in the building energy simulation domain, and there is an openly available direct conversion of CityGML as OWL [33]. From this snapshot of ontologies in the building domain, there is no shortage to satisfy a range of data modelling requirements for building and building related data. Therefore, methodologies for Linked Data generation to transform existing resources into Linked Data together with linking to authoritative building data, like that provided by the OSi, can provide a sound basis for interlinking these ontologies.

Several research projects have and are looking at the issue of linking Geospatial data with building data. The integration of IFC for the preconstruction stage of a building to support site planning, in terms of localization of materials and services appropriate for optimized productivity of a particular construction project, was examined in [34]. Other research has investigated the conversion of standards such as IFC directly into CityGML [35] [36] [37]. More recent work has specifically looked at converting IFC models into GeoSPARQL. GeoSPARQL is an Open Geospatial Consortium (OGC) standard which not only defines a vocabulary for representing geospatial data on the Semantic Web, but also specifies an extension to the SPARQL query language for processing that geospatial data [13].

None of this ongoing work was an exact fit for the OSi Prime2 geospatial data, but RDF and OWL enabled us to re-use and extend existing schemata. OGC's GeoSPARQL provides the core geospatial feature model and it was extended to create specific feature types like building, boundary, way from the Prime2 model. The W3C PROV-O ontology provided important primitive properties for describing the data sources, change authority and the series of changes to a feature like a building as it evolves over time. One thing that W3C PROV-O does not specify is the versioning strategy itself i.e. when and how new geospatial feature objects should be created and thus this is up to the users of PROV-O to decide. OWL provided ontology metadata such as authorship and versioning and the Dublin Core vocabulary [38] and the defacto metadata standard provided by the LODE (Live OWL Documentation) web service was an important guide.

2.4. Ordnance Survey Ireland Building Data

The Ordnance Survey Ireland (OSi) maintains GIS data for the Republic of Ireland. OSi aims to leverage user engagement with their geospatial information (and derived data and services), as it has a legal mandate to do so in Ireland. One of the initiatives they launched is called GeoHive (http://www.geohive.ie/), allowing easy access to publically available spatial data – but not as Linked Data (LD). Though OpenStreetMap, Google Maps, etc. allow people to easily engage with maps, the information provided by those are i) not authoritative, and ii) not always correct. One

[1] https://www.w3.org/TR/vocab-ssn/

of the major discrepancies that can be observed between these services and the information provided by the OSi are the points that refer to buildings. Where the former usually uses the entrance as the point, the OSi uses a building's centroid as a reference next to keeping track on which street the main entrance can be found. The latter can thus give a better indication of the size or location of a building with respect to the surrounding streets, for example.

The OSi aims to adopt the Linked Data principles as one means to publish its geospatial data. By doing so, it facilitates the exploration, adoption and use of OSi's authoritative geospatial datasets. In [39],[40], we reported on data.geohive.ie, which publishes and serves Ireland's authoritative boundary datasets as Linked Data on the Web. We note that the boundary data made available in the previous subsection was made available with an accessible license under OSi's open data release. Government departments and public-sector bodies under the National Mapping Agreement (NMA) [41] (an Irish agreement) have unrestricted access to most of OSi's geospatial data. With the NMA, one can request access to other datasets such as buildings and infrastructure. Stakeholders who fall under the NMA can request access to OSi data and services, which are also made available as dumps and shared over FTP.

Others who wish to avail of OSi's data that do not fall under this agreement, e.g., commercial entities, need to interact with the OSi's commercialization team and pay a license fee for obtaining the data, which can also be provided as dumps or web services. Nonetheless the OSi is in the process of making a subset of its building data publicly available without a license. This work addresses that subset. In the next sections we describe the creation of the ontology to represent that subset of building data which is being made publicly available.

3. Methodology

The methodology for generating the ontology from a data set and of converting the data and deploying it involves 5 high level steps presented in **Fig. 1**. **Step 1** consists of an analysis of the data set to identify suitable mappings to other available ontologies and data sets. The goal of this analysis is to (a) provide richer semantic descriptions of concepts and (b) to support interlinking of the data and further enrich the amount of available data to describe concepts. This first step requires a domain expert who is familiar with suitable ontologies and standards to identify suitable mappings. **Step 2** is the development of the ontology by an ontology engineer. This can be achieved using existing tools, such as Protégé [42] or through any text editor. **Step 3** is the process of writing the R2RML [43] mappings using a standard text editor, which are then stored as the RDF turtle syntax. R2RML supports converting relational databases to RDF datasets, or in fact, converting any tabular data (csv, tsv). This requires knowledge of R2RML and requires some knowledge of writing SQL queries.

To write meaningful mappings, the author of the R2RML must also understand the ontology and the structure of the data set. **Step 4** consists of running the R2RML Java processor developed and extended by the ADAPT Centre [44][45].

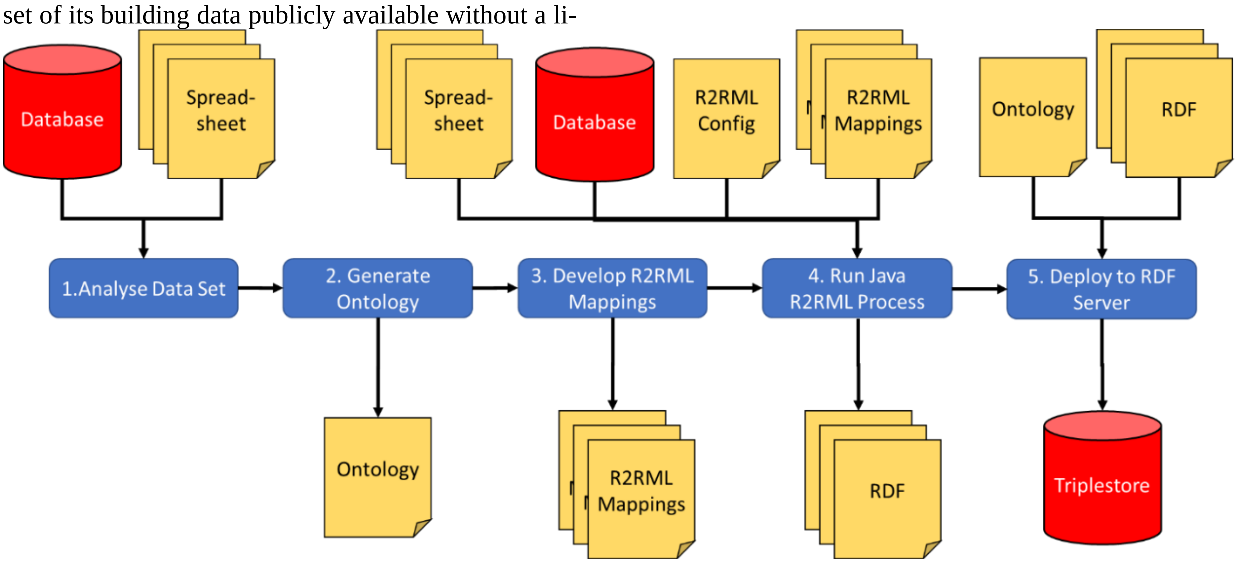

Fig. 1 Methodology for publishing RDF generated from tabular data

This is a manual process, which requires a small configuration file to specify the connection to the database, the name of the R2RML input file and the name of the RDF output file. **Step 5** is a manual process, which involves setting up an RDF triplestore and uploading the RDF, although this could be automated if required. If the data set is to support geometries, an appropriate server must be used, e.g. Parliament [46], Stardog [47], Strabon [48] or Fuseki [49]. Once the RDF is uploaded to the triplestore, SPARQL queries can be enacted over the available endpoint. Due to the restricted nature of the OSi building data, different licensing must be taken into consideration when accessing it. This is not the case for the building URIs, and these thereby provide a unique addressing scheme for linking, but for geometry and other building data an approach is described further in the paper for managing access control in this step. We now describe each step as applied to the OSi building data set in more detail.

4. Analysis of OSi building data Schema

This section describes the application of Step 1 and the analysis of the Prime2 DLM_CORE. DLM Core is a non-normalized subset of Prime2 provided to OSi customers and is the data which was provided to ADAPT for the purpose of this research. Most of the concepts carry over and it is expected that only small changes will have to be made to the mappings to use Prime2 directly in the future. It should be noted that due to intellectual property restraints on the internal structure of the DLM_CORE, we obfuscate the column names and include only exploratory examples of the R2RML mappings (section 6), and not the actual mappings. This section provides a description of the current understanding of the Prime2 building data as derived from the DLM Core data dump provided by the OSi. **Fig. 2** (an extension of the figure presented here [50]) depicts a high-level overview of the Prime2 Building related concepts, taken from the building data view in DML_CORE which has the point geometry data for the building, alongside all other concepts. The first we discuss here is an abstraction of the geometric point column, which we call *Placement*, related to the geolocation of the building. This is represented by a 2D point derived from determining the median of the polygon which represents the 2D footprint of the building. The second is taken from the 2D footprint (a polygon) called here *Geometry*. In Prime2, this may also be a 3D object when an additional height property is included, thus conforming to LOD1 in CityGML.

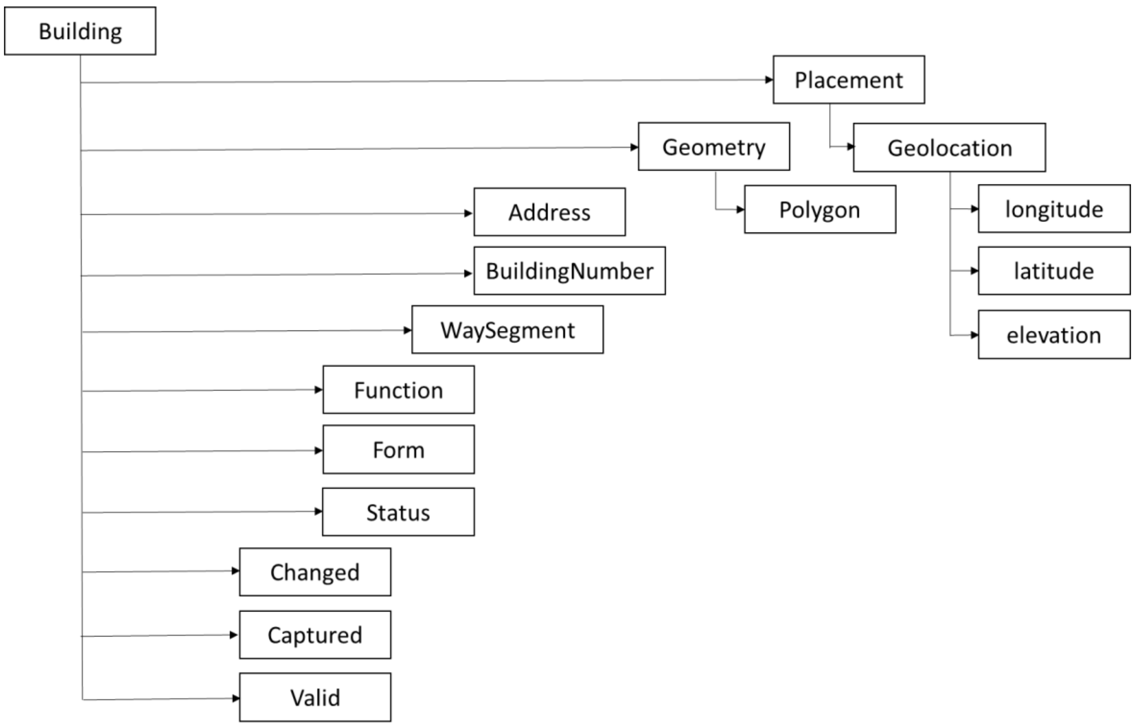

Fig. 2: Conceptual Overview of Prime2 (DLM_CORE) Building Data

Prime2 geometries are encoded as Oracle SDO geometric objects. These are a geometric description of a spatial object stored in a single row[2]. Each point of the polygon is represented by its own geodetic coordinate according to a reference system, for example Irish Transverse Mercator (ITM), or WGS84 [51]. There is also a value to indicate the resolution of the data, e.g. 20 meters, 50 meters or 100 meters, although these are not relevant for buildings, being used for boundary data, such as counties, electoral divisions, etc. The third concept is an Address. The Prime2 database references an address database called GeoDirectory[3] using a geo_id, which is an integer, and this is not yet required to be represented within the OSi.

The fourth and fifth concepts are building-specific. These are the *Form* and *Function* of the building. Form represents the physical form of the building, for example whether it is a "Building General" or "Barracks" and is an enumerated list of these type of values. Function, similarly to Form, is an enumerated list of values, but represents the use of the building, e.g. "Bank" or "Army Barracks". Some building forms have only one use, for example "Airport Terminal". Others, like building general can have multiple functions, e.g. airport building, bakery, courthouse, etc. It is also possible that a building will share a form and function, e.g. "Abbey". Form and function values exist for all geographic features, and two views are available in DLM_CORE for both form and function, which include both building and other non-building data, such as "Bog", or "Woodland Deciduous" for form or "Battlefield" for function.

Status represents the status of the building, e.g. 'Derelict', 'In Use', "Under Construction", "Proposed", "Dismantled", "In Ruin", "Site of" and "Disused". Status therefore has a relation to the life-cycle stage of the building. The *ZOrder* concept is for overlapping feature (e.g. bridges over rivers) and has one of three values a "1", "-1" or "0". The *Building-Number* is a number for the building (i.e. a house number "42"). Building *WaySegment* links the building entrance to a road or some other feature using a GUID. The referenced feature is not represented in an ontology and is not covered in this paper, only a reference to the GUID is maintained. The enumerated values, such as form, function and status, also have a direct numeric ID representation, are used to directly reference its value. The remaining concepts are related to who *Captured* the data e.g. the "OSi", how the data was *Changed* e.g. "Re-engineered", and who *Validated* this e.g. the "OSi". Captured, Changed and Valid are assigned to geometric objects, to the form, function, status, etc. In the next section, the representation of these concepts in the ontology is described.

5. OSi Building Ontology Development

This section presents Step 2, which has resulted in the OSi building ontology specification to support the DLM_CORE Prime2 building data in two graphs; one graph representing the current building data (i.e. most recent), and another graph (a provenance graph) which stores all changes made to the building data. The published ontology can be found here [52]. **Fig. 3** gives an overview of the ontology, visualized in the Protégé tool.

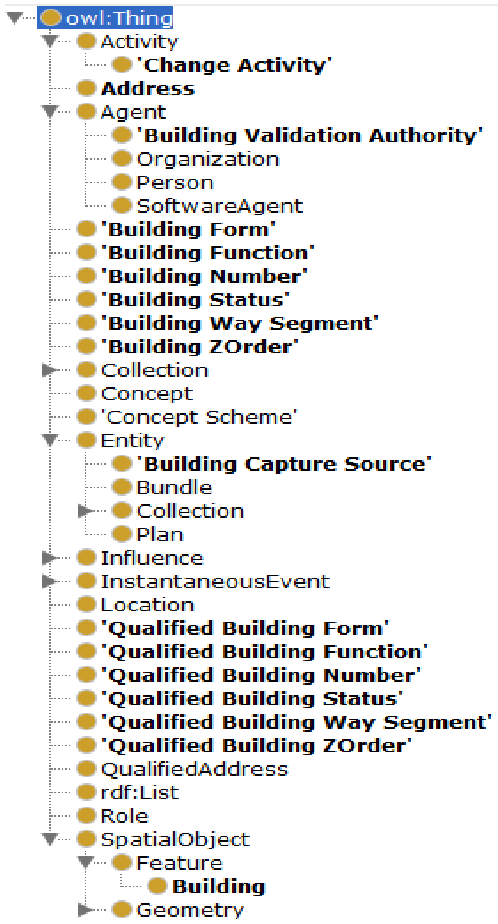

Fig. 3: OSi building ontology (viewed in Protégé)

[2] https://docs.Oracle.com/database/121/SPATL/sdo_geometry-object-type.htm

[3] https://www.geodirectory.ie/

Table 1 Used namespaces within this paper

Prefix	Namespace
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
osib	http://ontologies.geohive.ie/osib#
geo	http://www.opengis.net/ont/geosparql#
prov	http://www.w3.org/ns/prov#
acon	http://ontologies.geohive.ie/accesscontol#
xsd	http://www.w3.org/2001/XMLSchema#
dbp	http://dbpedia.org/ontology#

Before we continue, we first present a list of namespaces used throughout this paper. The ontology imports both GeoSPARQL [53] and the PROV-O ontologies [14] following W3C data on the web best practices for reuse of available vocabularies [54]. The *Building* concept in the ontology is represented in the same way as OSi boundary data has been represented e.g. *osib:Building* is a subclass of the *geo:Feature* from the GeoSPARQL ontology [39]. GeoSPARQL is used to capture the geospatial coordinates of the building, as well as the polygon shape of the floor print, using Well-Known-Text (WKT) representations of polygons. Buildings are represented by two geometries, their geolocation (WKT Point) and their footprint (WKT Polygon). The footprint is the *geo:defaultGeometry*. It should also be noted that Prime2 geometries are never more complex than Level of Detail LOD1 (in CityGML [55]), and these "3D" objects can be represented in WKT. For more complex building geometries (LOD2 and above), WKT may not be sufficient and alternative methods will need to be found. The Industry Foundation Classes (IFC) can represent complex geometries, and can be linked directly through the mapping process, e.g. through an object property relationship *osi:hasIfcOwlRepresentation*, but how appropriate these methods are for RDF-based geometry representations is an open research question [3], [8], [9]. For Prime2, it remains a fact that GeoSPARQL is sufficient to represent the available Prime2 geometries.

The OSi building ontology Form object property *osi:defaultForm* links it to its current enumerated Form value, *osi:hasQualifiedForm* which links it to the Building *osi:QualifiedForm* class and its object property *osi:hasForm* which inks it to a Form value, and the data property *osi:hasID* and an example enumeration value. *osi:QualifiedForm* is discussed in the later section on provenance. These classes and object and data properties are enough to express the building form as found in DLM_CORE. The same approach is used for Function and Status. The enumeration values for form, function and status are currently stored directly as instances in the ontology and 52 form, 145 function and 8 status values are represented, as these are currently referenced by existing building entries. When considering mappings to vocabularies for Form and Function, we need to consider other non-geospatial standards. For example, there is the US OmniClass [56] and UK Uniclass2 [57], both based on ISO 12006-2 classification for construction [58], which provide classifications for Form. Neither of these though appears to be as detailed as the OSi Form classification. IFC provides a means to model classification systems using IfcClassification, and could therefore provide additional semantics to the Form and Function concepts as defined in the OSi ontology through direct mappings to *IfcClassification*, e.g. using a *hasIfcClassification* property.

The BuildingSmart Data Dictionary (BsDD) [59] (BuildingSmart also developed the IFC standard) is an ISO 12006-3 based ontology for the building and construction industry, which includes tables from the Uniclass classification system. This rich vocabulary of terms could also satisfy Form and Function concepts as defined by OSi. It is therefore important to be able to support interlinking of OSi data with IFC, if BIM integration is to be achieved. Status as mentioned previously has a relationship to the buildings life-cycle. Life cycle stage classifications are, similarly to form and function, often defined at a national level. ISO 14040 defines life cycle assessment [60], and can potentially be used as a basis for building life cycle stages. Here *IfcClassification* may once again be an appropriate method for defining these classifications, where no other existing vocabularies exist. Address in Prime2 is currently only a single geo_id value which references the GeoDirectory data store. More complex addresses can be represented also using, for example, the *IfcPostalAddress* property.

5.1. OSi building data Provenance

Publishing Linked Data that describes the evolution of building geometries and other attributes (form, function, way segment, etc.) requires consideration of recording the changed attributes and recording metadata about the change event itself, e.g. who authorized the change? Collectively this information is known as provenance or lineage. There are two aspects of this data that must be included in the Linked Data model of DLM_CORE: For capturing the metadata on changes, we have used PROV-O, the W3C Recommendation for representing provenance, for handling the building Changed, Captured and Validation events recorded in DLM CORE (Prime2). Buildings are represented by two geometries, their geolocation (WKT Point) and their footprint (WKT Polygon). The footprint is the default geometry.

For recoding the historical values of changed attributes we must include them in an extended part of the building data model for Linked Data. In addition, the requirement to be able to track and store old versions of datatype properties is not directly supported by W3C PROV-O due to the resource-oriented nature of RDF models. Hence all datatypes must be modelled as qualified types (an object instead of a value) to enable additional context, e.g., relation date, to be attached to them in the RDF model. An alternative here would be to make use of the Ontology for Property Management (OPM) [61]. OPM, which extends PROV-O, supports modelling evolving, interdependent properties in knowledge graphs with a specific focus on Architecture, Construction and Engineering data. Future versions will examine OPM when aligning the OSi geospatial data with more complex building product data.

The current geospatial building data is stored in a default graph, provenance data is then stored in a separate provenance graph (see Fig. 4). In addition to the more complex qualified types e.g. the class *osib:QualifiedForm* and associated property *osib:hasQualifiedForm*, *osib:defaultForm* provides a simple accessor property which is included in the default graph and which corresponds to the current value of the property.

This follows the pattern established by the OGC in their GeoSPARQL ontology with both a *defaultGeometry* and *hasGeometry* property defined. All these aspects complicate the RDF (Linked Data) model and so they are stored in a second named graph (called "provenance") that complements the basic building information stored in the default graph. To support creation of metadata about changes to building attribute values, such as geometry, form, function etc. a class is created in the ontology called "Building Change" which represents a Change activity, which is a subclass of *prov:Activity*. In addition, for all the enumerated types related to buildings attributes or provenance metadata have instances created as part of the schema. Hence each 'form id' column value found in DLM CORE has a named instance created.

Fig. 4 illustrates the process of elaborating the buildings graph to add provenance information. We make a distinction in this figure between individuals and enumerated individuals, the latter of which are stored in the ontology. Using Form as an example, each building will have one or more *osib:hasQualifiedForm* relationships which point to a *osib:QualifiedForm* which points to a *osib:Form* value using *osib:hasForm* property. The *osib:QualifiedForm* also has the relationship *prov:generatedAtTime* which points to the 'change date' for that value and *prov:wasGeneratedBy* which points to a description of the provenance data.

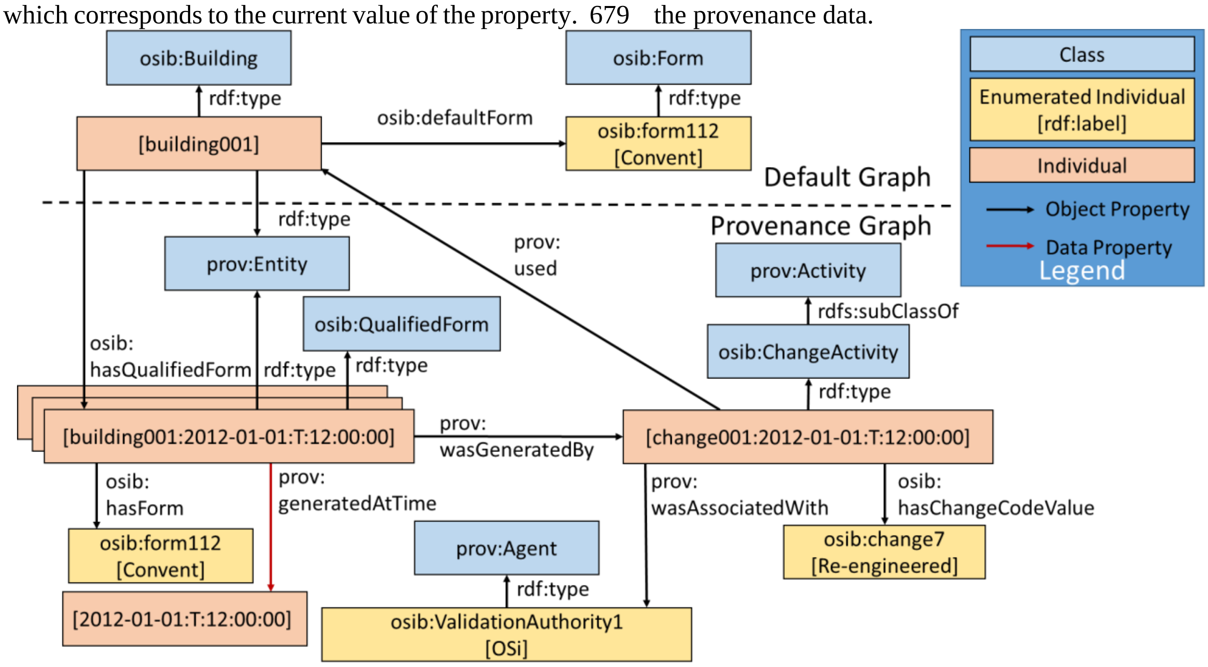

Fig. 4: Separation of Default and Provenance graph, here representing building Form

The provenance data then has the properties *prov:used* which points at the unique building id of the building the Form value belongs, or belonged, *osib:hasChangeCodeValue* which points to an enumerated list of Change Code Values (e.g. "Reengineered"), *osib:hasCaptureSourceValue* which points to an enumerated list of Capture Source Values (e.g. "OSi") and *prov:wasAssociatedWith* which point to an enumerated list of Validation Authorities (e.g. "OSi"). Finally, the provenance also uses again the *prov:generatedAtDate* which points to the change data.

6. R2RML Mappings for OSi building ontology

This section describes step 3, the development of the R2RML mappings, as applied to the DLM_CORE (Prime2) buildings and their evolution. These mappings enable the execution of step 4, the generation of RDF from the DLM_CORE data, which is stored in an Oracle 12c database, which includes building data describing the form, function, status, way segment, as well as the building geometry (both point and 2D polygon footprints). The mappings result in two graphs, a default graph which contains the most recent values for the building and a graph related to provenance data, i.e. a record of all the changes made to the building data. The mappings use the OSi building ontology to describe concepts, or alternatively direct mappings are specified in the R2RML mapping where these concepts are not explicitly included within the ontology, e.g. for provenance entities. The rationale and use of blank nodes and URNs are also discussed in relation to the representation of building geometries. The mappings are encoded in R2RML and consist of two parts:

- 1) A set of SQL queries to extract relevant building data from Oracle, and
- 2) A corresponding set of templates describing how the extracted data is transformed into RDF.

The SQL queries serve two purposes, a) to select the correct data from the correct tables in DLM_CORE and b) to filter the data so that we can create two different graphs. The way data is stored in DLM_CORE means that every concept (geometries, form, function, status, etc.) has a 'change date' column (which indicates if a change took place). All buildings which have a change to any concept, will have more than one row in the table. Therefore, all buildings with one row can be filtered and added to the default graph using the "Current Buildings without Changes" mapping. For data which has undergone a change, we must then use the "Current Buildings with Changes" mappings to filter only those with more than one row, and which have also undergone a change (this is done on a concept by concept mapping). Finally, for the provenance graph a set of "All Building Changes" mappings are needed. This uses a similar approach as for "Current Building with Changes" to filter all buildings with more than one row and return those concepts which have differences in their 'change date' values. For an overview of these three types of mappings, see **Table 1**. The next subsections describe these mappings and their results in more detail. Before though we briefly describe the rationale for using URNs.

6.1. Use of URNs

In the initial variations of the OSi Building ontology, geometries are handled using blank nodes. This was a decision made to prevent linking to the geometry of a feature rather than the feature itself. Blank nodes prevent this as they have no URI, but become an issue when managing provenance, because it becomes impossible for a provenance graph to point directly to a geometry if it is represented as a blank node. As it is necessary to store multiple geometric representations, the decision was made to move away from blank nodes to the use of URNs. An URN provides a unique identifier for a geometry, but does not allow an agent to obtain that resource with HTTP - hence it cannot be linked to as per Linked Data principles. A typical example of a URN is <urn:osi:geom:pnt:41c117b0a6e94625a33358724f9e8d9b1622007-04-26T00:00:00> which gives an indication of the type of geometry "pnt", its GUID "41c117b0a6e94625a33358724f9e8d9b162" and its creation date "2007-04-26T00:00:00".

It should be noted, that more recent developments around building geometry have seen the creation of two potential standards for managing representation of multiple geometries; the Ontology for Managing Geometry (OMG) and File Ontology for Geometry formats (FOG) ontologies, which [62], [63]. Future work will explore whether such approaches can be deployed to manage provenance in geometric data.

6.2. OSi Building Default Graph Mappings

This default graph only records the current or most up to date representation of the building (i.e. it omits the history of changes). Hence in this section we focus on those mappings to convert the DLM_CORE building data that represents the current (most recent entry) values for Buildings in the Oracle database, into RDF.

Table 1 An overview of the three mappings resulting in two graphs, with description

Graph and Mapping Name	Mapping Description
Default Graph: Current Buildings without Changes	This mapping converts all current buildings which only have at most one entry (i.e have undergone no changes)
Default Graph Current Buildings with Changes	These mappings (one for each value that can undergo a change e.g. Poly and Point Geometry, Function, Form, etc.) convert all buildings which have undergone a change for that particular concept by searching for all CHANGE_DATE with more than one unique value for a unique Building which has more than one entry/row in the data base. I.e. if Building 001 has more than one entry, the FORM mapping will search for all FORM_CHANGE_DATE for Building 001 and if there is more than one unique value, this has undergone a chance. It then returns the most recent Building row, based on that date.
Provenance Graph All Building Changes	These mappings (one for each value that can undergo a change e.g. Poly and Point Geometry, Function, Form, etc.) record all changes. I.e. if a concept has undergone a change, a provenance graph is created to capture a history of those changes based on the mappings.

Listing 1 provides a sample of the RDF output of the results of this mapping for the building with GUID 'babc6b00-bf32-4821-983a-5de8247c3367' with an RDF type of *geosparql:Feature* and *osib:Building*.

This RDF contains the current values for a buildings Form, Function, Status. The most recent form value is associated with building through the *osib:defaultForm* relationship as defined in the OSi building ontology. Each form value represents an instance defined directly in the ontology used to represent an enumerated list of values. This defines the extent of the descriptions for the form value. A similar mapping approach for building function and status is used. The WaySegment concept is given its own URI to enable extensions which will model WaySegments in a similar fashion to buildings. Currently, its URI points only to a string GUID for the WaySegment. Geometry, as described earlier, uses named graphs to represent the geometries, which have a unique URN based in the format.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix osib: <http://ontologies.geohive.ie/osib#>.
@prefix prov: <http://www.w3.org/ns/prov#>.
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.
@prefix geo: <http://www.opengis.net/ont/geosparql#>.
@prefix res: <http://data.geohive.ie/resource/building#>.

res:babc6b00-bf32-4821-983a-5de8247c3367 a osib:Building ;
  rdf:type geo:Feature ;
  osib:hasQualifiedForm res:form/babc6b00-bf32-4821-983a-5de8247c3367:2017-05-09T00:00:00Z,
                        res:form/babc6b00-bf32-4821-983a-5de8247c3367:2012-01-01T00:00:00Z .

res:form/babc6b00-bf32-4821-983a-5de8247c3367:2012-01-01T00:00:00Z>
  osib:hasForm osib:form55 ;
  prov:generatedAtTime "2012-01-01T12:00:00Z"^^xsd:dateTime ;
  prov:wasGeneratedBy res:prov/form/change/babc6b00-bf32-4821-983a-5de8247c3367:2012-01-01T00:00:00 .
```

Listing 1: Resulting RDF (turtle) from R2RML mapping for "Qualified Form", i.e. all Form values

```
@prefix osib: <http://ontologies.geohive.ie/osib#>.
@prefix prov: <http://www.w3.org/ns/prov#>.
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.
@prefix res: <http://data.geohive.ie/resource/building#>.

res:prov/form/change/babc6b00-bf32-4821-983a-5de8247c3367:2017-05-09T00:00:00Z
 osib:hasCaptureSourceValue osib:capture1 ;
 osib:hasChangeCodeValue osib:change7 ;
 prov:generatedAtTime "2017-05-09T00:00:00Z"^^xsd:dateTime ;
 prov:used osib:babc6b00-bf32-4821-983a-5de8247c3367 ;
 prov:wasAssociatedWith osib:validationAuthority1 .

res:prov/form/change/babc6b00-bf32-4821-983a-5de8247c3367:2012-01-01T00:00:00Z
 osib:hasCaptureSourceValue osib:Capture1 ;
 osib:hasChangeCodeValue osib:Change7 ;
 prov:generatedAtTime "2012-0101T00:00:00Z"^^xsd:dateTime ;
 prov:used osib:babc6b00-bf32-4821-983a-5de8247c3367 ;
 prov:wasAssociatedWith osib:ValidationAuthority1 .
```

Listing 2: Resulting RDF from R2RML mapping for Form PROV-O Change Activity.

That URN then is of type GeoSPARQL Geometry and has an asWKT representation of its geometry "POLYGON ((-6.16185506340804 53.4527688443413, -6.16180583892863 53.4527573549846, -6.16181520476239 53.4527430642236, -6.16175026781616 53.4527279085594, -6.16169339287773 53.4528147481574, -6.16180829989031 53.4528415664877, -6.16185506340804 53.4527688443413))". The mappings used for point geometry are used for polygon geometries with the exception that poly is represented using the *defaultGeometry* predicate. The "Current Buildings with Changes" mappings map the data in this same way. The difference here is that the queries must be run for every concept which can change over time, and for those concepts which have one or more 'change dates' with dissimilar values, all values are taken for the most recent 'change date' building. Next, we briefly describe some of the properties of these queries.

6.3. Oracle Queries for Managing Building Data

To enable data mapping within oracle, it was necessary to develop a suite of SQL queries. This section briefly describes the use of Oracle queries within the mappings. Listing 3 gives the R2RML mapping to query Oracle SDO geometries. These are converted into a WKT character object (CHAR) with a WGS84 coordinate system. In Oracle, this has a coordinate reference system id (EPSG[4] SRID) value of 4326[5]. Listing 4 gives the R2RML mapping to return all values for a concept (e.g. form) from a table (i.e. the table with the polygon footprint) that have undergone a change. In this example only those buildings which have two or more entries are selected using the group and count Oracle commands. Here we see the Oracle date converted into the format "YYYY-MM-DD" and appended with the string "T00:00:00Z" to the end so that it adheres to W3C Provenance data standard, as the required format is xsd:dateTime which mandates that a time be specified also.

[4] https://spatialreference.org/ref/epsg/

[5] https://epsg.io/4326

```
select
replace(GUID,'-', '')as GUID, replace(GUID,'-', '')as
GUID, TO_CHAR(SDO_UTIL.TO_WKTGEOME-
TRY(SDO_CS.TRANSFORM(SEGMENT, 4326))) AS
GEOM,
to_char(nvl(pnt_geom_change_date, '01-01-1921'),
'yyyy-mm-dd')||'T'||to_char(nvl(pnt_geom_change_date,
'01-01-1921'), 'hh24:mi:ss') as GEOM_CHANGE_DATE,
FORM_ID,
FUNC_ID
from TABLE_NAME where FORM_ID = 16 and
FUNC_ID = 632
```

Listing 3: A query to select and convert all geometries into WKT for a particular form and function

```
select
GUID,
(TO_CHAR(DATE, 'YYYY-MM-DD'),':T12:00:00') as
CONCAT_DATE, VALUE_ID
from (
    select * from TABLE_NAME
        where guid IN
            (select guid from TABLE_NAME
            group by guid having count(*) >1)
        Group by guid, TO_CHAR(DATE, 'YYYY-MM-
        DD') having count(*) <2
  )
```

Listing 4: A query to select all values of a concept (e.g. form) where a building has undergone a change (i.e. greater than one entry)

We have chosen the 12:00:00 hour time for all entries as no time data is provided. In the default graph the xsd:dateTime value is not needed, but is needed in the provenance graph, the mappings of which are described next.

6.4. *OSi Building Provenance Graph*

As discussed in section 5.1, the provenance data for buildings is recorded in a separate graph, called the provenance graph, which records when a building property value (e.g. Form, Function, Geometry) was changed, the type of change activity that was conducted, who captured the change and the validation authority who validated the change. Each *QualifiedForm* value is recorded using the *hasQualifiedForm* predicate, associating a *QualifiedForm* with a unique Building. The *QualifiedForm* then points to an enumerated Form instance, which records the value of Form. *QualifiedForm* is also associated with an OSi Change Activity, which is of type PROV-O Activity, and which is associated using the *prov:wasGeneratedBy*. The *QualifiedForm* also includes a *prov:generatedAtTime* value, which is taken from the 'change date' value for that concept.

The PROV-O Change Activity for capturing the Form value is associated to the Building it references by the *prov:used* predicate. The Change Activity records the 'change code' using the OSi building ontology *hasChangeCodeValue* predicate. This points directly at an enumerated value for recording the type of change activity. Similarly the *osib:hasCaptureSourceValue* records the 'capture source'. The validation authority is recorded using the *prov:wasAssociatedWith* predicate. Finally, the Change Activity also includes a *prov:generatedAtTime* value, which again is taken from the concatenated 'change date' value for that concept. Listing 4 shows the results of the mapping for Building GUID 'babc6b00-bf32-4821-983a-5de8247c3367'. As can be seen, there are two Form entries, recorded here as *QualifiedForm generatedAtTime* '2012-01-01T12:00:00Z' and '2017-05-09T12:00:00Z'. These both point to a PROV-O Change Activity as shown in Listing 4, and records that they have *hasChangeCodeValue* Change7, *hasCaptureSourceValue* Capture1 and *wasAssociatedWith* validationAuthority1. The enumerated values are available in the OSi building ontology are "Reengineered", "OSi" and "OSi" respectively.

6.5. *Generating the RDF*

Once the ontology and mappings have been defined, and the RDF generated, the next step (**step 4**) is to execute an openly available R2RML processor[6] which reads in the R2RML file, and connects to the Oracle database, and queries the data, converting it to RDF. This is all managed through the definition of a simple configuration file, which specifies inputs, outputs, and the source of the data (which can be csv files, or a relational database). The resulting RDF is uploaded to a triplestore. In the next sections we describe the deployment (relating to **step 5**) and then services for accessing the data.

7. OSi building data Deployment and Access Control

The ontology is deployed on the OSi servers[7], and a subset of the Prime2 RDF is available publicly[8]. To store the RDF locally, there are several triplestores

[6] https://opengogs.adaptcentre.ie/debruync/r2rml
[7] http://ontologies.geohive.ie/building/osib
[8] http://data.geohive.ie/downloadAndQuery.html

that currently support geospatial functions, e.g. Strabon and Fuseki (both difficult to set up), or Stardog (requires a license for geospatial support), Parliament is recommended due to ease of set up and its open license. Using the Parliament web client, the RDF is uploaded to the triplestore, SPARQL queries can then be enacted over the available endpoint. To support geospatial functions, the RDF must be indexed first. The building data is not made available publicly, but instead relies on a licensing agreement to enable access. This access is governed by an access control approach.

We developed a new access control approach for geospatial Linked Data and developed a prototype implementation of this approach. The purpose of this access control approach and implementation is to ensure that customers can only access the portions of the building data that they have purchased a license for and ensure they cannot access data they have not paid for. This approach and implementation consist of five main components which are discussed individually below: (i) an Access Control Model which is used to model licenses and templates; (ii) a RESTful API; (iii) a Template Selector component; (iv) a Template Analyzer component; (v) a Query Processor component. The prototype implementation is a web application, implemented in Python, hosted on an Apache web server. This communicates with the Parliament triple store hosting the OSi building data. The licenses and templates that are created as part of our access control approach are also stored in separate named graphs in the triple store. A high-level architecture diagram of the access control implementation is presented in **Fig. 5** below [64].

The access control model is an OWL vocabulary that was developed to model both licenses and templates as part of our approach. The license section of the model is used to capture the details a customer will be able to access. It also captures information such as when the license is due to expire and how many times it may be used in conjunction with a template to retrieve data. In addition, a license also captures which existing templates are permissible to be used with it. In **Fig. 6** below we display an overview of the license portion of the access control model.

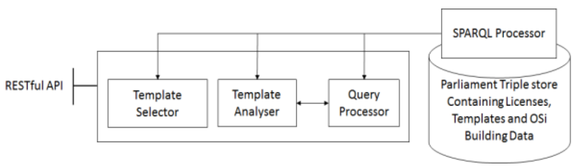

Fig. 5: High Level Architecture of our Access Control Implementation

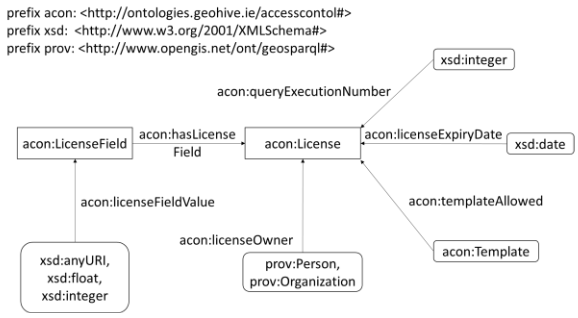

Fig. 6: License Portion of the Access Control Model

A license instance will be of class *acon:License*. A license is linked to an owner via the *acon:licenseOwner* property. The *acon:queryExecutionNumber* property indicates the remaining number of times a license can be used to execute a template. The *acon:licenseExpiryDate* property indicates the license expiry date. The *acon:templateAllowed* property is used to link the templates that are allowed to be used with a license (which is automatically added by the Template Selector component). A license can have multiple *acon:LicenseFields* and these can have multiple *acon:licenseFieldValues*. The access control model contains multiple sub-classes of the *acon:LicenseField* class, specialized for geospatial data which are used to explicitly capture what a license field specifies.

The template section of the access control model is concerned with modelling how data can be accessed. This is done through linking a SPARQL query, which contains place holder variables, to a template. The template explicitly models each placeholder variable, specifying the range of values that a variable can contain. **Fig. 7** displays an overview of the template section of the access control model. A template instance will be of class *acon:Template*. The *acon:templateDescription* property is used to link a description of a templates functionality. The *acon:query* property is used to link a SPARQL query to a template.

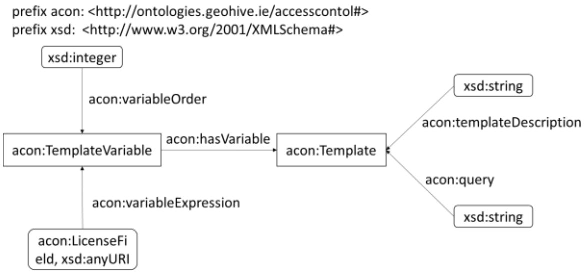

Fig. 7: Template Portion of the Access Control Model

The SPARQL query is represented as a string and the placeholders within the query take the form of *$variable1*, *$variable2* and so on.

The *acon:TemplateVariable* class is used to model the variables within a query and are linked to a template instance via the *acon:hasVariable* property. Variables have an order number, linked via the *acon:variableOrder* property. The *acon:variableExpression* property is used to link the values that are allowed for a variable placeholder, to a variable.

7.1. RESTful API

Access to data using our access control approach is achieved through a RESTful API call. We support two types of call: a 'status' call and a 'query' call. The purpose of the status call is so a user can check their status with regard to valid licenses they possess and to find out which templates are usable with each license (as well as a description and basic instructions on the use of the templates). The status call will invoke the template processor component. In a status call, a user must specify their user ID: http://localhost/acon/status/{userID}.

The purpose of a query call is for a user to access data. The query call will invoke the template analyzer and query processor components. In a query call a user must specify their user ID, the license they wish to use, the template they wish to use and the variable values for the template. This way the user accessing the data does not need to know the SPARQL query language to access the data. An example of a query is shown in **Listing 5**.

```
http://localhost/acon/query/{userID}/{LicenseID}/{TemplateID}?variable1={variable_1_value}&variable2={variable_2_value}&variableN={variable_N_value}
```

Listing 5. Sample SPARQL query

7.2. Template Selector Component

The template selector component discovers which of the existing templates are allowed to be used by a user, based on the licenses they possess. This is done through analyzing a user's licenses and any existing templates. The license fields are checked against each variable of each template to see if a license field is allowable. If a license field is allowable for each variable in a template, then that template can be used with that license. The template selector component also ensures that a license has not expired and it has at least one query execution left. When the analysis of the template selector is complete, it creates a link between licenses and templates via the *acon:templateAllowed* property. In addition, the template selector component also returns a message to the user, which indicates which templates can be used for each of their license(s), the description for a template and a description of how to use each template (i.e. what can be specified for a templates variables).

7.3. Template Analyzer Component

The template analyzer component validates a query call. This component ensures that the details specified in a query call are correct according to the values allowable by the variables of the template specified and also allowable according to the user's license specified. If any of the details specified in the query call are not allowable, then an error is returned to the user and the implementation proceeds no further. If all the details specified in the query call are allowable, then the query processor component is next invoked.

7.4. Query Processor Component

The query processor component is invoked by the template analyzer component after it has checked and validated all the details specified in a query call are allowable. The query processor component retrieves the SPARQL query associated with the template specified in the query call. From there it substitutes the variable values (specified in the query call) into the place holders in the SPARQL query. The query processor component then sends the query for execution to the SPARQL processor, retrieves the results and return them to the user. Once the data is available, it can be explored how to interlink it with other building data sources to enrich the data provided by OSi, such as an IFC file, DBpedia data and open governmental data.

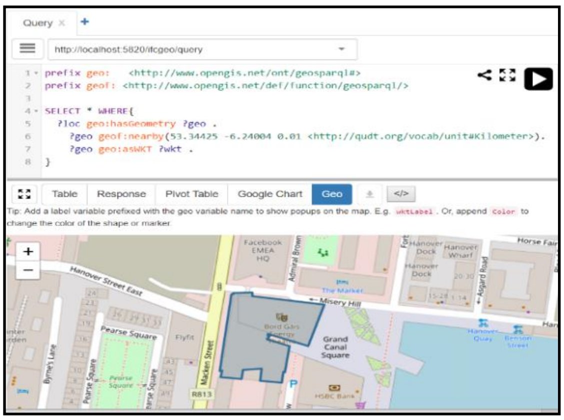

Fig. 8 GeoSPARQL to locate features near DBpedias “Board Gáis Energy Theater” in Dublin

8. Interlinking Building Data

In this section three use cases (**Fig. 9**) are presented which make use of GeoSPARQL to support the interlinking of different data sets which contain building information with OSi building data.

8.1. Interlinking with DBpedia data.

In [65] an example of linking OSi building data with DBpedia data has been demonstrated through the use of a SPARQL construct query (**Listing 6**) for adding GeoSPARQL WKT geometries to DBpedia buildings, and then running a ‘nearby’ Geospatial function (this function is supported by the Stardog [47] triplestore) to return all buildings within a certain geographic area. shows the result of this query run on YASGUI [66]. This data integration supports the enrichment of OSi building data, which includes a polygon footprint, form and function with other data available on DBpedia, such as the architect of the building (see **Fig. 8**), its address, opening date, seating capacity, etc. Even in this simple example, which only looked at one building, it can be seen that this type of integration can support building data enrichment.

```
PREFIX ge: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX geo: http://www.opengis.net/ont/geosparql#
PREFIX dbp: http://dbpedia.org/ontology/

CONSTRUCT {
?s geo:hasGeometry [geos:asWKT ?point ] .
}
WHERE
{
?s a dbo:Place.
?s
dbo:locationCountry <http://dbpedia.org/resource/Republic_of_Ireland>.
?s dbp:latd ?lat .
?s dbp:longd ?long .
bind( STRDT(concat("POINT(", ?lat , " ", ?long, ")"),
geo:wktLiteral) as ?point)
}
```

Listing 6. Query to map DBpedia to GeoSPARQL

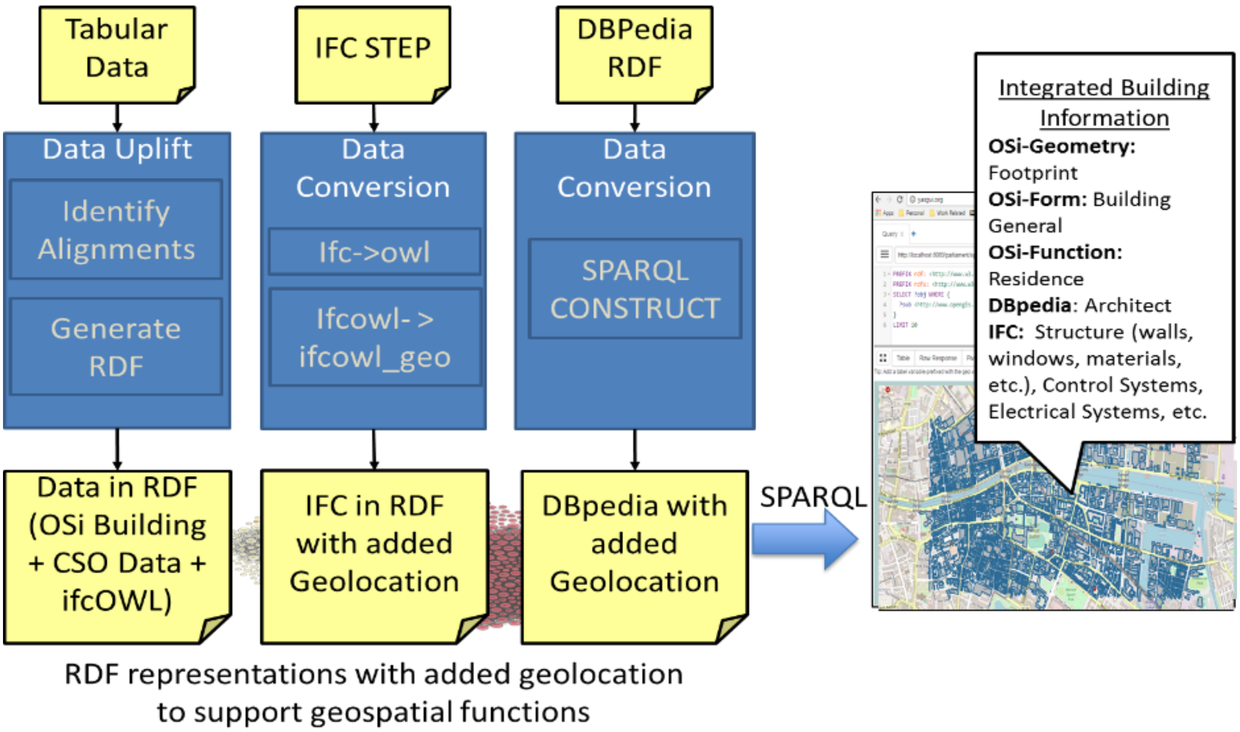

Fig. 9: Uplift of building data into GeoSPARQL to support interlinking

8.2. Interlinking with IFC data.

More recent work examined specific issues related to integration of IFC data with geospatial data, such as that provided by the OSi, and the particular issues related to integration of different geometric representations between the geospatial and BIM domains [8]. In [50] and [9], the conversion of ifcOWL geometries into GeoSPARQL, as well as other non-geospatial geometries, such as OBJ files, has been developed. This approach demonstrated the alignment of IFC geometry with its geospatial equivalent, enabling the geotagging of all building elements which have a geometric representation in the IFC model. The challenge remains with IFC that the Longitude and Latitude is not aligned with the origin of the Cartesian coordinate system used to represent the buildings geometry. Methods for aligning two WKT geometries are therefore being explored.

Fig. 10 gives an example of the use of a GeoSPARQL function geo:sfOverlaps to match geometries that overlap. This has the potential to support building matching with the combination of additional properties, that is, where two buildings do not align correctly. Using the additional footprint area, a SPARQL filter function returns results based on the filter condition. Two filter conditions specify that the two returned building WKT geometries overlap and that the calculated area of one WKT polygon is within +/- 25% the area of the overlapping polygon. Both aligned polygons, representing the same physical building, are returned if these two filter conditions are met. The two WKT geometries are taken from an ifcOWL model and another from a Building Topology Ontology (BOT) model [67].

Once the data is available, it can be interlinked it with other building data sources to enrich the data provided by OSi, such as an IFC file, DBpedia data and open governmental data. By combining OSi data and IFC, a much richer set of building data becomes available for querying. This approach also supports more complex queries over multiple data sets, for example DBpedia. Ultimately, the goal is to develop a method for querying large data sets of building and building related data, based upon the authoritative geometries and URIs which are available through the OSi.

This makes it possible to execute queries which return subsets of open data, along with licensing information for data which is considered to have some commercial value. Current work is examining the conversion of IFC geometries of floors and spaces into WKT, a representation supported by GeoSPARQL, thus enabling the full power of geospatial functions for SPARQL queries, such as nearby, within, touches, etc. which can support queries related to energy efficiency, navigation, building control, etc. This work must take into consideration the complex relationships with IFC to support the geometric representation of these components [9].

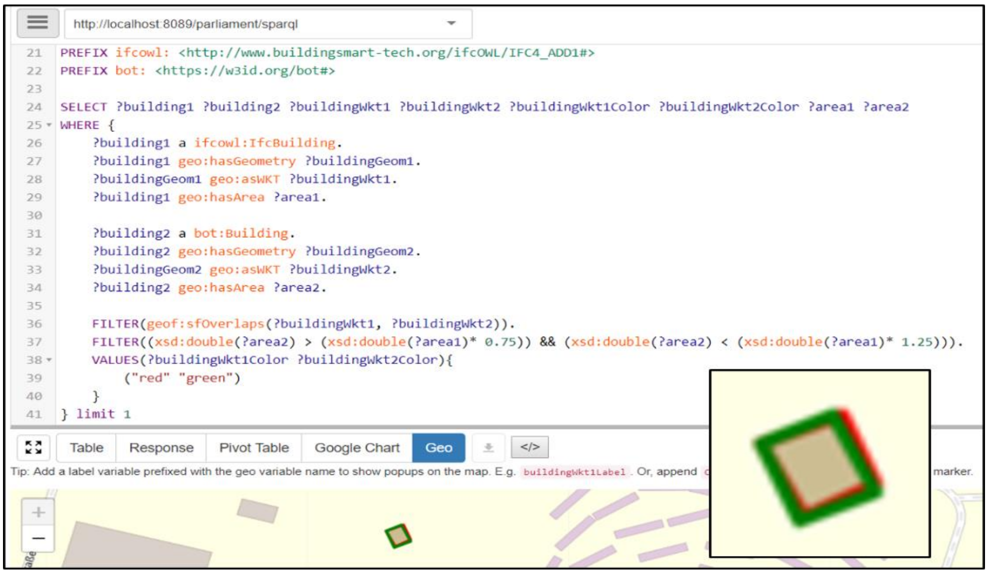

Fig. 10: SPARQL to detect possible alignments between overlapping building geometries. Screenshot is taken from YASGUI, the Geo results page draws the ifcOWL (red) and BOT (green) geometries.

8.3. Interlinking with Irish open data.

As part of Ireland's open data strategy for 2017-2022[9], the Irish government has launched data.gov.ie which has over 8,800 open data sets published. These can be downloaded in a range of formats, including JSON, CSV and XML. Some of these data sets also have geospatial coordinates provided. In [10] three data.gov.ie open data sets were converted into RDF using R2RML. Next a WebApp called "GViz" was developed to support the querying of these data sets using configurable GeoSPARQL queries through a WebGL & Google Maps interface. This work demonstrated that simple to use interfaces, built upon Web of Data technologies, can provide integrated data on buildings, such as data from data.gov.ie and OSi, which can be queried by non-expert users. Such interfaces can provide a basis for querying and integrating data sets, through processes such as the alignment process described in [9].

In [68], the OSi authoritative geometry building data set was interlinked with the Irish Central Statistics Office (CSO)[10] data. This work mainly explored mappings to datasets using the RDF Data Cube Vocabulary [69], a W3C standardized vocabulary for the publication of multi-dimensional (statistical) datasets. This paper presented the development of the OSi Prime2 building ontology and the methodology for generating the OSi building RDF for both the default (current) building values, and historical (change) values.

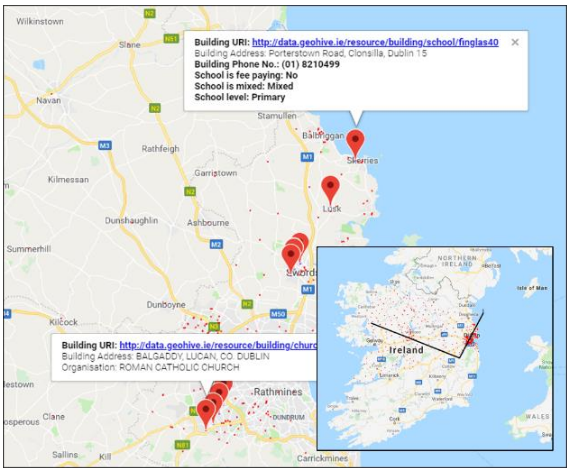

Fig. 11: Polygon area based selection using GViz

The geometric data is published using GeoSPARQL, which also supports additional geospatial functions within SPARQL queries. The addition of this geospatial building data allows one to analyze, link, explore, and even build data analysis applications on top of several datasets using Semantic Web technologies. In the following section, we discuss some of our major findings with respect to the approach taken, and challenges met.

9. Discussion

In general, the Linked Data tools and approach has proven itself to meet the requirements for the publishing of geospatial data. Despite this, when building the OSi ontology and deployment workflow the following limitations of the Linked Data approach were found:

1. Complexity of ifcOWL models. The component geometry-based models of IFC are very powerful but increase complexity beyond the level appropriate for typical Linked Data users.
2. Dominance of WGS84 co-ordinate system in non-specialist web data standards. The OSi data natively uses the ITM (Irish Transverse Mercator) co-ordinate system and since this is supported by GeoSPARQL's WKT encoding it is possible to publish the Linked Data as ITM but for typical Linked Data consumers this is likely to be a barrier to adoption due to the dominance of WGS84 as a baseline in non-specialist international Web standards such as those by the W3C. The impact of this remains to be seen. Linked Data based on the GeoSPARQL model does enable both representations to be published simultaneously but at a great cost in storage and potentially bandwidth (given the verbose WKT representation).
3. Tool immaturity can still cause development bottlenecks. Some platforms we used natively support advanced Linked Data features such as R2RML but most deployments have not stretched them to national scale use cases and so ongoing vendor support is essential to avoid development delays.
4. Political and commercial considerations around open data. Since Linked Data is strongly associated with the Open Data movement it can create communication challenges around the deployment of Linked Data systems, especially where commercially sensitive data is

[9] https://www.gov.ie/pdf/6572/?page=1

[10] https://www.cso.ie/en/index.html

concerned. Our access control solution was developed in partial response to this concern.

10. Conclusion and Future Work

The R2RML mappings presented in this paper are developed for managing OSi building data and are based upon the OSi ontology, GeoSPARQL and the W3C PROV-O ontology. The resulting data sets should satisfy current capabilities for querying both the present building values, their geometry using geospatial functions, as well as historical values based upon selection criteria, such as the date, who captured the data, the types of capture activity and who validated the data. This can be achieved for all concepts modelled within the ontology, such as Geometry (point and poly), Function, Form, Status, Address, WaySegment, ZOrder and Building Number. As not all OSi building data is currently open, and depends on a license, the paper also presented a method for querying data based on licensing information. This allows queries which hide sensitive or restricted building data.

The feasibility of the approach has been demonstrated through the integration of OSi building data with, DBpedia data, IFC data and Central Statistics Office data. To further support access to building data, we believe it is of key importance to create this authoritative basis for developing integrated data on buildings, which can be linked to. OSi can provide just such a data hub, and OSi URIs for buildings and features can become the central hub in the Rep. of Ireland for integrating BIM data.

Future work will explore linking the data with some OWL models to further validate the approach and initial work has been begun in this respect [9]. Also being explored are the use of flat geometries to describe building floor plans (on a storey by storey basis), which can then be published and shared to support indoor navigation, control and energy management. Of interest also is the integration of more complex geometries, as GeoSPARQL currently does not support 3D geometries and parallel work is exploring how multiple geometric representations and coordination systems can be maintained for buildings. Finally, work is underway to create a more generic ontology to manage form and function for geospatial and building data and which builds on existing classification systems such as OmniClass and Uniclass, called Geoff.

Acknowledgements

This research was funded with the financial support of SFI ADAPT Centre and Ordnance Survey Ireland. The ADAPT SFI Centre for Digital Media Technology is funded by Science Foundation Ireland through the SFI Research Centres Programme and is co-funded under the European Regional Development Fund (ERDF) through Grant # 13/RC/2106.

References

[1] L. O. Oyedele *et al.*, "Big Data in the construction industry: A review of present status, opportunities, and future trends," *Adv. Eng. Informatics*, vol. 30, no. 3, pp. 500–521, Aug. 2016.

[2] Y. Chen, "Industrial information integration—A literature review 2006–2015," *J. Ind. Inf. Integr.*, vol. 2, pp. 30–64, Jun. 2016.

[3] C. Eastman, P. Teicholz, R. Sacks, and K. Liston, *BIM Handbook: A guide to Building Information Modeling for owners, managers, designers, engineers and contractors*. 2011.

[4] A. Borrmann, M. König, C. Koch, and J. Beetz, *Building Information Modeling: Technology Foundations and Industry Practice*, 10.1007/97. Springer, 2018.

[5] P. Pauwels, S. Zhang, and Y. C. Lee, "Semantic web technologies in AEC industry: A literature overview," *Automation in Construction*, vol. 73. pp. 145–165, Jan-2017.

[6] K. Afsari, C. Eastman, and D. Shelden, "Cloud-Based BIM Data Transmission: Current Status and Challenges," in *Proceedings of the 33rd International Symposium on Automation and Robotics in Construction (ISARC)*, 2017, pp. 1073–1080.

[7] S. Azhar *et al.*, "Building Information Modelling (BIM) uptake: Clear benefits, understanding its implementation, risks and challenges," *Renew. Sustain. Energy Rev.*, vol. 75, pp. 1046–1053, Dec. 2016.

[8] K. McGlinn, A. Wagner, P. Pauwels, P. Bonsma, P. Kelly, and D. O'Sullivan, "Interlinking geospatial and building geometry with existing and developing standards on the web," *Autom. Constr.*, vol. 103, pp. 235–250, 2019.

[9] J. O'Donovan, D. O'Sulllivan, and K. McGlinn, "A method for converting IFC geometric data into GeoSPARQL," in *Proceedings of the 7th Linked Data in Architecture and Construction Workshop*, 2019, pp. 7–20.

[10] K. McGlinn, D. Blake, and D. O'Sullivan, "GViz - An Interactive WebApp to Support GeoSPARQL over Integrated Building Information," in *Companion Proceedings of The 2019 World Wide Web Conference on - WWW '19*, 2019, pp. 904–912.

[11] Ordnance Survey Ireland, "Prime2 - Data Concepts & Data Model Overview," p. 17, 2014.

[12] T. Heath and C. Bizer, "Linked Data: Evolving the Web into a Global Data Space," *Synth. Lect. Semant. Web Theory Technol.*, vol. 1, no. 1, pp. 1–136, Feb. 2011.

[13] X. Lopez, "GeoSPARQL - A geographic query language for RDF data A proposal for an OGC Draft Candidate Standard," 2012.

[14] T. Lebo, S. Sahoo, D. McGuinness, K. Belhajjame, J. Cheney, and others, "PROV-O: The PROV Ontology. W3C Recommendation, 30 April 2013," *World Wide Web*

Consort., 2013.

[15] ISO, "ISO 16739:2013 Industry Foundation Classes (IFC) for Data Sharing in the Construction and Facility Management Industries," *International Organization for Standardization (ISO)*, 2013. [Online]. Available: http://www.iso.org/iso/catalogue_detail.htm?csnumber=51622. [Accessed: 22-Jan-2019].

[16] "DBpedia," 2019. [Online]. Available: http://wiki.dbpedia.org/. [Accessed: 22-Jan-2019].

[17] "Databases - CSO - Central Statistics Office," 2019. [Online]. Available: http://www.cso.ie/en/databases/. [Accessed: 22-Jan-2019].

[18] X. Liu, R. Liu, G. Wright, J. Cheng, X. Wang, and X. Li, "A State-of-the-Art Review on the Integration of Building Information Modeling (BIM) and Geographic Information System (GIS)," *ISPRS Int. J. Geo-Information*, vol. 6, no. 2, p. 53, Feb. 2017.

[19] C. Tang, L. Da Xu, and S. Feng, "An agent-based geographical information system," *Knowledge-Based Syst.*, vol. 14, no. 5–6, pp. 233–242, Aug. 2001.

[20] G. Gröger and L. Plümer, "CityGML – Interoperable semantic 3D city models," *ISPRS J. Photogramm. Remote Sens.*, vol. 71, pp. 12–33, 2012.

[21] J. E. Taylor and P. G. Bernstein, "Paradigm Trajectories of Building Information Modeling Practice in Project Networks," *J. Manag. Eng.*, vol. 25, no. 2, pp. 69–76, Apr. 2009.

[22] B. McAuley, A. Hore, and R. West, "BIM Innovation Capability Programme - Irish BIM Study," *Articles*, 2016.

[23] A. Galiano Garrigós, T. Kouider, and J. Scott, "The Big BIM battle: BIM adoption in the UK for large and small companies," in *Healthy Buildings: Innovation, Design & Technology*, 2016, pp. 53–66.

[24] B. McAuley, A. Hore, and R. West, "BICP Global BIM Study - Lessons for Ireland's BIM Programme," Jan. 2017.

[25] buildingSMART, "BuildingSMART Home Page," 2014. [Online]. Available: http://www.buildingsmart.org/.

[26] R. Howard and B.-C. Björk, "Building information modelling – Experts' views on standardisation and industry deployment," *Adv. Eng. Informatics*, vol. 22, no. 2, pp. 271–280, Apr. 2008.

[27] McGraw Hill Construction, *Smart Market Report: The Business Value of BIM in North America: Multi-Year Trend Analysis and User Ratings (2007-2012)*. 2012.

[28] W. Terkaj and A. Šojić, "Ontology-based representation of IFC EXPRESS rules: An enhancement of the ifcOWL ontology," *Autom. Constr.*, vol. 57, pp. 188–201, Sep. 2015.

[29] "ifcOWL ontology," 2019. [Online]. Available: http://ifcowl.openbimstandards.org/IFC4_ADD1/index.html. [Accessed: 22-Jan-2019].

[30] L. Daniele, F. den Hartog, and J. Roes, "Created in close interaction with the industry: the smart appliances reference (SAREF) ontology," *Int. Work. Form.*, 2015.

[31] D. Bonino and F. Corno, "DogOnt - Ontology Modeling for Intelligent Domotic Environments," Springer, Berlin, Heidelberg, 2008, pp. 790–803.

[32] M. Kofler and W. Kastner, "Towards an ontology representing building physics parameters for increased energy efficiency in smart home operation," *Proc. 2nd Cent. Eur.*, 2013.

[33] ISI, "A direct translation of the CityGML XMLSchema to OWL." [Online]. Available: http://cui.unige.ch/isi/onto/. [Accessed: 22-May-2017].

[34] E. P. Karan and J. Irizarry, "Extending BIM interoperability to preconstruction operations using geospatial analyses and semantic web services," *Autom. Constr.*, vol. 53, pp. 1–12, 2015.

[35] S. Donkers, H. Ledoux, J. Zhao, and J. Stoter, "Automatic conversion of IFC datasets to geometrically and semantically correct CityGML LOD3 buildings," *Trans. GIS*, 2015.

[36] Y. Deng, J. C. P. Cheng, and C. Anumba, "Mapping between BIM and 3D GIS in different levels of detail using schema mediation and instance comparison," *Autom. Constr.*, vol. 67, pp. 1–21, 2016.

[37] L. Zhu and H. Mao, "Building Interior Model Interpolation between IFC and CityGML," *Open Constr. Build. Technol.*, 2015.

[38] D. C. M. I. DCMI, "DCMI: Dublin Core Metadata Element Set, Version 1.1: Reference Description," 2012. [Online]. Available: http://www.dublincore.org/specifications/dublin-core/dces/. [Accessed: 04-Jul-2019].

[39] C. Debruyne, É. Clinton, L. McNerney, A. Nautiyal, and D. O'Sullivan, "Serving Ireland's geospatial information as linked data," in *CEUR Workshop Proceedings*, 2016, vol. 1690.

[40] C. Debruyne, K. McGlinn, L. McNerney, and D. O'Sullivan, "A lightweight approach to explore, enrich and use data with a geospatial dimension with semantic web technologies," in *Proceedings of the Fourth International ACM Workshop on Managing and Mining Enriched Geo-Spatial Data - GeoRich '17*, 2017, pp. 1–6.

[41] Ordnance Survey Ireland, "National Mapping Agreement," 2017. [Online]. Available: https://www.osi.ie/services/national-mapping-agreement/. [Accessed: 22-Jan-2018].

[42] "The Protégé Ontology Editor and Knowledge Acquisition System," 2019. [Online]. Available: http://protege.stanford.edu/. [Accessed: 22-Jan-2019].

[43] A. J. Nathan and A. Scobell, "How China sees America," *Foreign Affairs*, vol. 91, no. 5. pp. 1–34, 2012.

[44] "r2rml," 2019. [Online]. Available: https://opengogs.adaptcentre.ie/debruync/r2rml. [Accessed: 22-Jan-2019].

[45] C. Debruyne and D. O'Sullivan, "R2RML-F: Towards Sharing and Executing Domain Logic in R2RML Mappings," *Proc. Work. Linked Data Web, LDOW 2016, co-located with 25th Int. World Wide Web Conf. (WWW 2016)*, 2016.

[46] Parliament, "Parliament High-Performance Triple Store," 2017. [Online]. Available: http://parliament.semwebcentral.org/. [Accessed: 22-Jan-2018].

[47] Stardog Union, "Stardog—The Knowledge Graph Platform for the Enterprise," 2017. [Online]. Available: https://www.stardog.com/. [Accessed: 22-Jan-2018].

[48] "Strabon," in *The Semantic Web*, 2008, pp. 289–307.

[49] "Fuseki: serving RDF data over HTTP," 2019. [Online]. Available: https://jena.apache.org/documentation/fuseki2/. [Accessed: 22-Jan-2019].

[50] K. McGlinn, C. Debruyne, L. McNerney, and D. O'Sullivan, "Integrating Ireland's Geospatial Information to Provide Authoritative Building Information Models," in *In 13th International Conference on Semantic Systems*, 2017.

[51] "GNSS Reference Systems," 2019. [Online]. Available: https://www.osi.ie/resources/reference-information-2/gnss-reference-systems/. [Accessed: 22-Jan-2019].

[52] "OSi Building Ontology," 2019. [Online]. Available: http://ontologies.geohive.ie/osi/building/index.html. [Accessed: 22-Jan-2019].

[53] O. G. Consortium and others, "OGC GeoSPARQL-A geographic query language for RDF data," *OGC Candidate Implement. Stand.*, vol. 2, 2012.

[54] "W3C, Data on the Web Best Practices," 2017. [Online]. Available: https://www.w3.org/TR/dwbp/.

[55] "CityGML," 2019. [Online]. Available: https://www.citygml.org/. [Accessed: 22-Jan-2019].

[56] "OmniClass," 2019. [Online]. Available: http://www.omniclass.org/. [Accessed: 22-Jan-2019].

[57] "Uniclass2," 2015. [Online]. Available: http://www.cpic.org.uk/uniclass/. [Accessed: 22-Jan-2019].

[58] "ISO 12006-2:2015," 2019. [Online]. Available: https://www.iso.org/standard/61753.html. [Accessed: 22-Jan-2019].

[59] "buildingSMART Data Dictionary Browser," 2018. [Online]. Available: http://bsdd.buildingsmart.org/. [Accessed: 22-Jan-2019].

[60] ISO, "ISO 14040:2006," 2006.

[61] M. H. Rasmussen, M. Lefrançois, P. Pauwels, C. A. Hviid, and J. Karlshøj, "Managing interrelated project information in AEC Knowledge Graphs," *Autom. Constr.*, vol. 108, p. 102956, Dec. 2019.

[62] A. Wagner, M. Bonduel, P. Pauwels, and U. Rüppel, "Representing construction-related geometry in a semantic web context: A review of approaches," *Autom. Constr.*, vol. 115, p. 103130, Jul. 2020.

[63] A. Wagner, M. Bonduel, P. Pauwels, and R. Uwe, "Relating geometry descriptions to its derivatives on the web," in *Proceedings of the 2019 European Conference for Computing in Construction*, 2019, vol. 1, pp. 304–313.

[64] A. Meehan, K. Fatema, R. Brennan, É. Clinton, L. McNerney, and D. O'Sullivan, "License and template access control for geospatial linked data," in *CEUR Workshop Proceedings*, 2018, vol. 2110, pp. 22–33.

[65] K. McGlinn, C. Debruyne, L. McNerney, and D. O'Sullivan, "Integrating Building Information Models with Authoritative Irish Geospatial Information," in *In 16th International Semantics Web Conference*, 2017.

[66] L. Rietveld and R. Hoekstra, "YASGUI: Not just another SPARQL client?," in *CEUR Workshop Proceedings*, 2013, vol. 1056, pp. 1–9.

[67] M. H. Rasmussen and P. Pauwels, "Building Topology Ontology." [Online]. Available: http://www.student.dtu.dk/~mhoras/bot/index-en.html. [Accessed: 22-May-2017].

[68] K. McGlinn, C. Debruyne, L. McNerney, and D. O'Sullivan, "Interlinking Ireland's Authoritative Geospatial Information On Buildings With Central Statistics Office Data," in *10th Annual European Forum for Geography & Statistics*, 2017.

[69] Richard Cyganiak and Dave Reynolds, "The RDF Data Cube Vocabulary," *W3C Recommendation*, 2014. [Online]. Available: https://www.w3.org/TR/vocab-data-cube/. [Accessed: 12-Jun-2019].