

Silvennoinen, Heidi et al.

Article

A semantic web approach to land use regulations in urban planning: The OntoZoning ontology of zones, land uses and programmes for Singapore

Journal of Urban Management

Provided in Cooperation with:

Chinese Association of Urban Management (CAUM), Taipei

Suggested Citation: Silvennoinen, Heidi et al. (2023) : A semantic web approach to land use regulations in urban planning: The OntoZoning ontology of zones, land uses and programmes for Singapore, Journal of Urban Management, ISSN 2226-5856, Elsevier, Amsterdam, Vol. 12, Iss. 2, pp. 151-167,
<https://doi.org/10.1016/j.jum.2023.02.002>

This Version is available at:

<https://hdl.handle.net/10419/271498>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Research Article

A semantic web approach to land use regulations in urban planning: The OntoZoning ontology of zones, land uses and programmes for Singapore

Heidi Silvennoinen^{a,*}, Arkadiusz Chadzynski^b, Feroz Farazi^b, Ayda Grišiūtė^a, Zhongming Shi^a, Aurel von Richthofen^{a,c}, Stephen Cairns^a, Markus Kraft^{b,d,e}, Martin Raubal^f, Pieter Herthogs^{a,**}

^a Singapore-ETH Centre, Future Cities Lab Global Programme, CREATE Campus, 1 Create Way, #06-01 CREATE Tower, Singapore, 138602

^b CARES Cambridge Centre for Advanced Research and Education in Singapore, 1 Create Way CREATE Tower, #05-05, Singapore, 138602

^c Arup, Joachimsthaler Str. 41, Berlin, 10623, Germany

^d Department of Chemical Engineering and Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge, CB3 0AS, United Kingdom

^e School of Chemical and Biomedical Engineering Nanyang Technological University, 62 Nanyang Drive, Singapore, 637459

^f Institute of Cartography and Geoinformation, ETH Zürich, Stefano-Franscini-Platz 5, CH, 8093, Zürich, Switzerland

ARTICLE INFO

Keywords:

City planning

Master plan

Zoning

Knowledge graph

Web 3.0

ABSTRACT

Semantic web technologies have the potential to significantly improve urban regulatory data access, integration and usability, with potentially large implications for planning practice. Ontologies are a cornerstone of the semantic web. In this paper, we describe OntoZoning, an ontology representing relationships between zoning types, land uses and programmes (more specific land uses) in Singapore. We link the ontology to geospatial data stored in a knowledge graph, which allows executing multi-domain queries on urban data. We demonstrate how such a semantic web based approach can improve access to and usability of land use regulation data, and in particular facilitate site selection and exploration. We also discuss the difficulty of defining some concepts in the land use regulation field, and how OntoZoning could be linked to a broader semantic-web based urban planning regulatory framework.

1. Introduction

Semantic web technologies have potentially wide-reaching benefits for the production of, access to and use of urban planning information. The overarching vision of the semantic web – sometimes also termed ‘Web 3.0’ – is to link individual pieces of data and facts together in a machine-readable way, rather than linking documents. Instead of using search words to find relevant documents or websites, the goal is to retrieve data directly, regardless of which database, document, website, or even format they are stored in (Patel & Jain, 2019). These technologies promise to increase the efficiency, scope and accuracy of information searches, and therefore improve the decision making of a wide range of organisations (Rudman et al., 2016). The potential benefits are especially noticeable when accessing and using geographic data, and it has been suggested that semantic web technologies may even induce paradigm shifts in Geographic Information Science (GIScience) and urban planning (Kuhn et al., 2014; Mai et al., 2020; Potts, 2020).

* Corresponding author.

** Corresponding author.

E-mail addresses: heidi.silvennoinen@ibi.baug.ethz.ch (H. Silvennoinen), herthogs@arch.ethz.ch (P. Herthogs).

This paper explores how the semantic web technology paradigm could be applied in the urban planning field, by creating OntoZoning, an ontology of land use regulations in Singapore, and testing its applications, particularly for site selection and exploration. An ontology is a key part of the semantic web technology stack which enables computers to understand the meaning of and relationships between different kinds of information (Rudman et al., 2016). The present work is part of the Cities Knowledge Graph project, a broader research effort to support urban planning through improved data representation, access and evaluation by using knowledge graphs, triple stores and autonomous software agents (Chadzynski et al., 2021, 2022a; Grisiute et al., 2022a, 2022b). This broader infrastructure allows evaluating the usefulness and practical implications of the OntoZoning ontology as part of a semantic web technology stack. In particular, it allows linking OntoZoning to geospatial plot data, and then conducting automated queries of what could be built where, replacing a manual process of searching for this information within many different documents and files.

Improving the accessibility and usability of land use regulations in this way is relevant from the point of view of urban planning theory. Urban planning theory is increasingly influenced by big data analytics, but there is also a recognition that cities are complex and unpredictable, and that planning should therefore be a continual collaborative process between many stakeholders (Portugali, 2012; Potts, 2020; Kandt and Batty, 2021). By improving stakeholders' access to land use regulatory information, ontologies could facilitate the collaborative process. Using ontologies to link regulatory information to further information supports data-driven urban models and evaluations of planning policies and scenarios (Heath, 2010; Potts, 2020).

The paper adds insights on an under-researched topic at the intersection of land use planning regulations and semantic web technologies. Most work on semantic web technologies in the urban sphere relates to non-regulatory data, for example, integrating building information and city information models, or categorising existing land uses and land covers without taking into account the needs of planners (von Richthofen et al., 2022; Xue et al., 2021; Ahlqvist et al., 2016; Gribb and Czerniak, 2016, pp. 1–19). Meanwhile, research on digitising urban planning regulations has typically not used semantic web approaches (e.g. Walczak, 2021; Zhang and Schnabel, 2017). Existing work at the intersection of urban regulations and semantic web technologies has linked map instances to documents (e.g. Winkels, 2007), defined allowed land uses based on intersections between regions (Hoekstra et al., 2010), semantically annotated land use regulations on the web (Iwaniak et al., 2016), and created ontologies of land use regulations (Chichkova et al., 2020; Montenegro et al., 2012).

The present paper offers three main contributions to the existing literature on land use and semantic web technologies. The first contribution relates to data accessibility and achieving the linked data vision. In contrast to work linking urban planning documents to maps (e.g. Winkels, 2007), or annotating existing regulatory documents (Iwaniak et al., 2016), the present work enables queries of the specific statements that documents contain, on a ready-made knowledge graph. Secondly, this paper provides a blueprint for integrating a wider range of more granular land use regulations, in addition to representing general guidelines on which land uses are generally allowed in the same region (Hoekstra et al., 2010). This is an important consideration for achieving accurate query results of allowed land uses, given that context-specific exceptions to land use regulations are common (Hirt, 2012). Finally, the present work differs from existing land use ontologies (Chichkova et al., 2020; Katsumi & Fox, 2018; Montenegro et al., 2012) through its integration with the broader knowledge graph infrastructure which allows links to geospatial data and visualisations of query results.

This paper is structured as follows. Section 2 begins by providing necessary background information on ontologies, Singapore's land use regulation system, and the definition of the key term 'land use.' Section 3 explains how the ontology was built. Section 4 contains the ontology itself and use-cases related to multi-criteria site selection through queries. The benefits and drawbacks of the presented semantic web technology approach are evaluated in Section 5, while Section 6 contains conclusions and future research outlooks.

2. Background

2.1. Role of ontologies in the semantic web

An ontology is defined as a shared conceptualisation of a field of interest that formally specifies the meanings of and relationships between different concepts (Gruber, 1993). A conceptualisation is a simplified, abstracted view of the field of interest. A specification fixes the language used to describe the field, including its concepts, relationships and attributes. While such specifications implicitly exist in any database and even in natural language, an ontology makes them explicit and formal, and thus machine-readable (Guarino et al., 2009). Ontologies are commonly represented in description logics and serialised in OWL (Web Ontology Language, see McGuinness et al., 2004).

Ontologies play a key role in the semantic web framework (Jacob, 2003). Achieving the goal of linking (and thus being able to query) individual facts or entities that are related requires computers to understand the links between these facts or entities. In conventional relational databases, meaning is separate from the structure of the database, and is often described in separate database schema or data dictionaries (Kuhn et al., 2014). In the semantic web approach, data are stored as subject-predicate-object triples whose contents follow the structure of the ontology, i.e. the representation of how the concepts in the domain relate to each other. Thus understanding the structure of the data follows from understanding the concepts and their relations in the domain itself. Ontologies therefore determine how computers classify data into concepts, and what kinds of relations between different data/concepts they can find.

Ontologies and semantic web technologies also facilitate the integration of data from different sources or domains, as different ontologies can be easily linked (Trojahn et al., 2022). In a semantic web approach, all domain data are stored as triples, which means that linking knowledge and data from different domains only requires aligning the meanings of the shared concepts in the ontology from each domain.

In contrast to linking triples based on ontologies, integrating relational databases would require aligning the structure *and* semantics of the data (see Kuhn et al., 2014, for a detailed explanation and examples). This may require complicated scripts that are not able to

function if the underlying databases are changed, and discrepancies and inconsistencies may occur if different terms are used in different ways in the different datasets. In the semantic web approach, the simple triple structure remains constant and easy to query and understand, even as the knowledge graph expands. Triple statements and queries are also translatable into natural language (Höffner et al., 2017). Each subject, object and predicate is a term with a unique identifier (a Uniform Resource Identifier or URI) in an ontology, ensuring that each term has a unique meaning and thus lessening scope for misunderstanding.

The potential benefits of the semantic web approach are especially noticeable when accessing and using geographic data, which often require complex data models, and combine data from many different formats and sources (Kuhn et al., 2014; Mai et al., 2020). For example in the urban planning domain, evaluating the impacts of land use planning may require investigating the links between land uses and energy consumption, infrastructure and mobility, urban heat island effect, or the total quantity of housing city-wide (Engin et al., 2020; Levine et al., 2019; Mohamed et al., 2019; Tan et al., 2020; Williams et al., 2021; Yazdanie & Orehounig, 2021).

2.2. Urban planning regulations in Singapore

Singapore's planning regulations are an interesting test case for the ability of ontologies and semantic web technologies to integrate different datasets and facilitate targeted queries across datasets. Like several European planning regulatory systems surveyed by Hirt (2012), Singapore's regulations operate at different scales, ranging from the national-level regulations to street block-level regulations, and come in different formats from different sources.

The main regulatory document is the nationwide Master Plan, a map that brings together regulatory urban data from different departments (Urban Redevelopment Authority, 2019b). Most importantly, it states each plots' zoning type and Gross Plot Ratio (GPR), as shown in Fig. 1. An associated text document (the Master Plan Written Statement, Urban Redevelopment Authority, 2019c) provides an overview of each zoning type in terms of its allowed land uses and their maximum or minimum use quantum (i.e. the allowed quantity of the land use in terms of gross floor area), and also specifies how to interpret zoning type and GPR rules in exceptional cases, such as when the plot is partly covered by a waterbody. In addition, separate online Development handbooks (e.g. Urban Redevelopment Authority, 2019a) contain more detailed lists of land uses allowed in each zoning type.

In addition, Singapore's planning regulatory system contains many region or even plot-specific regulations which specify and override the general land use and intensity regulations of the Masterplan (Urban Redevelopment Authority, 2019c). These may regulate building mass through e.g. required road setbacks or building heights, but also allowed land uses. For example, even though a plot may have the 'Residential' zoning type, the plot's specific location may influence what kinds of residential programmes are actually allowed there. If the plot is in a 'Landed housing area', then condominiums, flats and serviced apartments are not allowed there, even though these are generally allowed in residential plots. Likewise, serviced apartments are only allowed in 'Residential' zone plots that are located at the 'fringe' of a residential neighbourhood, and not within the neighbourhood. Although the OntoZoning ontology does not represent these local regulations, it has been designed to accommodate location- or type-specific exceptions, as illustrated in Section 4.6. Ongoing work in the Cities Knowledge Graph project is incorporating the local regulations more fully.

Singapore government agencies have also recognised the benefits of facilitating access to urban regulatory data more generally. The Singapore Land Authority (SLA) has worked on a project to automate cadastral job processing, which involves building a semantic 3D model of Singapore (Soon & Khoo, 2017). Additionally, as part of an effort to enhance planning with the use of Artificial Intelligence (AI), the URA is building a knowledge base which semantically links a wide range of planning-related data (Huang, 2022; Urban Redevelopment Authority, 2020). The goal is to help urban planners and industry stakeholders access relevant data more efficiently, enable urban analyses and 'evaluat[ions] of planning options', 'automat[e] routine tasks', and develop models to assess land use needs based on factors such as demographics (Huang, 2022).

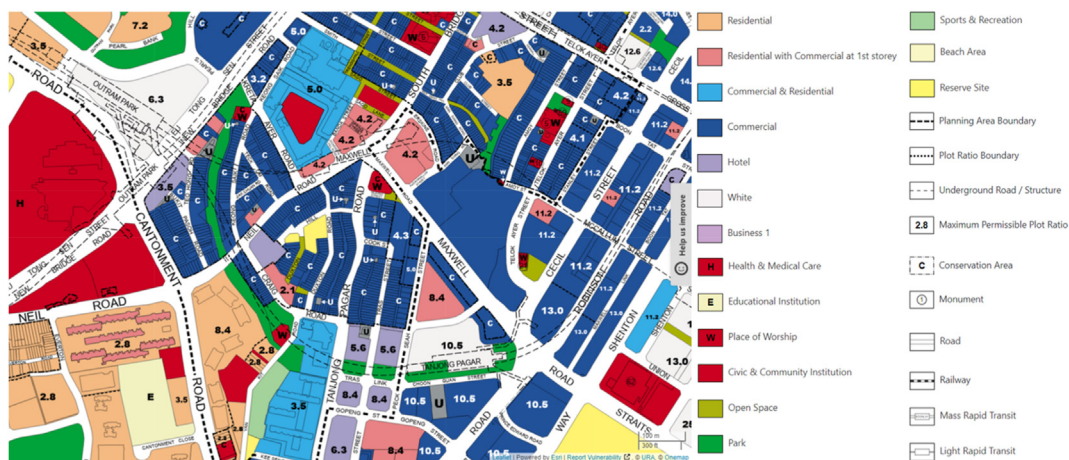


Fig. 1. An extract of the Singapore 2019 Master Plan, as shown in URA Space (Urban Redevelopment Authority, 2019b). Only zoning types and symbols visible in the map excerpt are included in the legend.

2.3. Defining key terms

Given that ontologies rely on precise definitions of terms and are intended to be reused in different contexts, it is necessary to consider what exactly is meant by ‘zoning’ and ‘land use’, in general and in Singapore specifically. ‘Zoning’ may be generally defined as ‘allocating different activities or uses ... to specific units of area within a region’ (Stewart et al., 2004). This is also how the term is used in Singapore, as discussed in Section 2.2. Different zoning systems place different levels of emphasis on restricting land use. For example, Euclidean zoning focuses primarily on separating land uses while performance zoning regulates mainly measurable impacts such as noise levels, almost regardless of which land uses causes them (Marwedel, 1998). However, a comparison of zoning and planning regulations in the United States and Europe by Hirt (2012) showed that all systems have at least some regulations governing which land uses are allowed where. ‘Land use’ is therefore another central concept in the field of planning regulations and is typically defined as a human activity that takes place within space (Bibby & Shepherd, 2000; Dickinson & Shaw, 1977). It is thus distinct from land cover, which describes the physical matter existing on the Earth’s surface (Ahlqvist et al., 2016).

However, classifying land according to its use is surprisingly difficult. Firstly, each activity may consist of a wide range of tasks with different levels of specificity; thus the activities chosen to describe land use are partially subjective and influenced by the interests of the classifier (Bibby & Shepherd, 2000). Secondly, activities may be associated with specific configurations of physical matter and types of ownership, which may also be of interest (Dickinson & Shaw, 1977). Arguably, this is the case in the zoning domain. Zoning documents usually do not describe allowable land uses purely from the perspective of activities, but may also specify whether the land is used by a public or private institution, and the kind of physical form that may be built on the land. Particularly physical structures are often strongly associated with activities: for example, the intension of the term ‘school’ includes various activities related to education. As discussed above in Section 2.2., documents by Singapore’s Urban Redevelopment Authority also uses the term ‘land use’ in this way, variously using it to refer to general concepts like ‘commercial use’, physical developments at different scales which at times imply ownership, like ‘mall’ or ‘private medical clinic’ and activities such as ‘shopping’ or ‘entertainment’. Against this background, Singapore’s land use regulations do not seem atypical in their use of the terms ‘zoning’ and ‘land use,’ which has relevance for the reusability of the OntoZoning ontology.

3. Methods

Following the ontology creation methodology by Mizen et al. (2005), we built our ontology in six steps (see Appendix A.1 for details). First, we defined the purpose and scope of the ontology with the help of competency questions (Section 3.1). Second, we reviewed existing ontologies for reuse potential (Section 3.2). Third, we collected the relevant source material and listed and defined terms to be included in the ontology (Section 3.3). Fourth, we translated the terms found into networks of concepts, relationships and attributes, recording any information lost and evaluating the network against the purpose and competency questions (Section 3.4). Fifth, we created and evaluated the logical ontology using Protégé (Section 3.5). Finally, we linked the ontology to geospatial data to provide a foundation for querying which plots allow which land uses, and, ultimately, the plots’ zoning regulations more generally (Section 3.6).

3.1. Purpose and scope of the OntoZoning ontology

When creating an ontology, the first step should be to define its purpose (Mizen et al., 2005). We did so by formulating a set of competency questions, essentially queries that the ontology should be capable of representing and answering (Katsumi & Fox, 2020). Competency questions enable checking that the ontology contains correct and sufficient concepts and relationships to represent the domain of interest.

As stated in the Introduction, the primary purpose of the ontology is to represent Singapore’s land use regulations in a way that allows expansion to other planning regulations later. This purpose was specified through the following competency questions.

- CQ1. What kinds of land uses, or their mixes, are allowed on a given plot?
- CQ2. Where are all the plots that allow a specific land use, or mix of land uses?
- CQ3. What minimum and maximum use quantum for specific land uses exist for different zoning types?
- CQ4. Where can I find a plot that allows a particular gross floor area of a specific land use to be built?

Further competency questions that were identified as important to take into account, but are beyond the scope of the present paper, include.

- CQ5. What kinds of land uses are allowed on a given plot given its zoning type, but also given its location in an area with additional requirements?
- CQ6. What is a typical quantity and distribution of uses for a plot with a certain zoning type and a certain GPR?

The additional competency questions (CQ5-6) require integrating the ontology and plot data with further data. In the case of CQ5, this means further regulatory data, including on geospatial data on areas with specific regulatory requirements. CQ6 implies the use of dynamic big urban data sources, such as government datasets, Google Maps Places or Open Street Maps (OSM) tags (Shi et al., 2022). These data sources would ideally be dynamic because land uses at the more specific level of Google Places or OSM tags can change frequently. Interoperability with such datasets would enable advanced uses, such as simulating and analysing the effects of planning on

e.g. urban energy systems, mobility patterns, or economics models (von Richthofen et al., 2022). Thus, throughout the ontology design process, we considered extendibility to other regulations and urban data.

3.2. Existing ontologies' potential for reuse

The reusability of existing ontologies was reviewed based on the objectives and competency questions described above. Although the OntoZoning ontology is focused on Singapore (with its own unique zoning types), other ontologies might contain reusable land use classes and their relations to zoning types. Six relevant, existing ontologies were analysed from the perspective of reuse (Barramou et al., 2020; Chichkova et al., 2020; Codescu et al., 2011; Katsumi & Fox, 2019; Katsumi & Fox, 2020; Kuo & Hong, 2012; Montenegro et al., 2012). The first four ontologies were not reused due to limitations in the number of useful concepts included, or differing focus (see Appendix A.2 for an analysis of each one). Below is a summary of the two interrelated ontologies that were chosen for reuse.

The first of these is the LBCS-OWL2 ontology (Montenegro et al., 2012), designed to enrich geospatial parcel data with land use data. The ultimate goal is to improve access to urban data (including e.g. population data), to support the formulation of design solutions. The land use ontology describes land based on five key dimensions defined by the American Planning Association's Land Based Classification Standards model (LBCS): 'Function', 'Activity', 'Structure', 'Site development status', and 'Property rights.' This ontology has significant potential for reuse, given the methodological approach to selecting classes based on the LBCS, the high number of relevant activity classes, and the similarity of its purpose compared to the present work. This ontology was not available online and could thus not be reused; however, an amended version that is available was reused (see next paragraph).

The iCity suite of ontologies (Katsumi & Fox, 2019; Katsumi & Fox, 2020) is proposed as a standard for transportation planning, to support integration and reuse of data. The suite contains an LBCSv2 ontology to describe land use (this is an extended version of the LBCS ontology described above). It also has separate ontologies to represent agents, transport infrastructure and several foundational concepts related to transportation planning, such as time.

Of the five land use dimensions in the LBCSv2 ontology, activities are most relevant for present purposes, considering the meaning of the term 'land use' in the URA documents (see Section 2.3). An activity is defined as 'an observable characteristic of land based on actual

S/No	Zoning	Uses	Examples Of Developments	Remarks
4	Commercial	These are areas used or intended to be used mainly for commercial development. Recreation Clubs may be allowed subject to evaluation by the competent authority.	Developments for: 1. Offices 2. Mixed Uses (e.g. Office/ Shopping /Cinema/Hotel/Flat) 3. Convention/Exhibition Centre 4. Commercial School 5. Bank 6. Market/Food Centre/Restaurant 7. Cinema 8. Entertainment 9. Foreign Trade Mission/Chancery	The developments in this zone are subject to controls on the type and quantum of commercial uses as determined by the competent authority.
5	Hotel	These are areas used or intended to be used mainly for hotel development.	1. Hotel 2. Backpackers' Hostel	At least 60% of the total floor area shall be used for hotel room floors and hotel related uses as defined in the Planning (Development Charges) Rules. Commercial and residential uses may be considered by the competent authority subject to control on the use quantum as determined by the competent authority and they shall not exceed 40% of the total floor area.

Fig. 2. An extract of the Master Plan Written Statement (Urban Redevelopment Authority, 2019c). The yellow and pink highlights draw attention to how this document contains more general 'uses' and more specific 'developments', which however encompass not only physical developments at different scales but also activities, general concepts like 'Entertainment', and sometimes imply ownership. The green highlights illustrate the discretion left to the URA. The parts highlighted in blue contain examples of use quantum regulations. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

use [...]. It describes what actually takes place in physical or observable terms (e.g., farming, shopping, manufacturing, vehicular movement, etc.). An office activity, for example, refers only to the physical activity on the premises, which could apply equally to a law firm, a non-profit institution, a court house, a corporate office, or any other office use.'

3.3. Collecting terms to include in the ontology

When collecting terms to be included in the ontology, it seemed natural to start with the Singapore regulatory data: the Master Plan, its associated Written Statement and the URA's Development handbooks for individual zoning types. The Master Plan data contains each plot in Singapore, along with the assigned zoning type and GPR in KML format (see Section 1). The Written Statement (Fig. 2) and Development handbooks describe the meaning of each zoning type in terms of the 'uses' and 'developments' that are, may, or may not be allowed. Here, we explain how we translated the land use terms found in the URA documents into two concepts, 'Land use' and 'Programme', in the ontology, and how we supplemented the list of terms with programmes from two other sources, Google Maps and the LBCSv2 ontology.

The first challenge in collecting and organising terms from the URA documents was defining land use, given the variety of concepts that the term encompasses in the documents (see e.g. Fig. 2). As discussed in Section 2.3., existing literature was helpful in arriving at an initial definition of land use as an activity or a physical development which may imply a certain form of ownership.

A second challenge was finding a suitable level of detail of land uses to include in the ontology. The different URA documents describe land use at different levels of detail, from more general concepts like 'Commercial uses' to specific concepts like 'Bar' or 'Hotel rooms' or even 'Hotel meeting room' in the Development Handbooks (Urban Redevelopment Authority, 2019a). One benefit of more general terms is that they make it easy to convey the character of a zoning type without listing all specific land uses that are allowed. On the other hand, the more specific land uses help to clarify the meaning of the general terms, especially in the case of more ambiguous terms like 'community institution' and 'civic institution.' Therefore, one requirement for the ontology was to include both general and specific land use terms.

The third step was to consider how to arrange the land use terms in a way that allows easily extending the ontology with further URA regulations, such as use quantum, and regulations from other contexts or countries. The maximum and minimum use quantum in the Master Plan (discussed in Section 2.2. and also highlighted in Fig. 2) generally apply to a single general land use concept such as 'Commercial uses' or 'Commercial and residential uses', which can encompass many specific land uses like 'Restaurant' or 'Bank.' Therefore, including use quantum in the ontology requires linking general land uses to more specific land uses.

The most obvious solution to this would be to form a hierarchy of land uses, however this is problematic for two reasons. First, more general land uses allowed in different zoning types might not actually have identical meanings. For example, the zoning type 'Commercial with Residential at First Storey' allows commercial uses except those that are likely to cause disamenity to residents, such as nightclubs, while commercial uses allowed in the zoning type 'Commercial' also include nightclubs. Thus there can be no single 'Commercial' land use that encompasses a single set of more specific land uses. Secondly, having a hierarchy is problematic because different zoning types allow different quantities of the same general land use. For example, at least 60% of GPR in 'Commercial' zones must be allocated to 'Commercial uses', while Hotel zones have a maximum use quantum of 40% for 'Commercial uses.' It was deemed important to accommodate for variation in allowed sets of land uses in a zoning type, and for different use quantum per land use in different zoning types, especially since Hirt (2012) suggests that similar regulations are also common outside of Singapore.

For these reasons, land uses in OntoZoning were ultimately split into two different concepts: 'Land use' and 'Programme.' These roughly correspond to the respective columns 'Land use' and 'Development type' in the Master Plan Written Statement (see Fig. 2). Programmes are defined as activities or physical developments that may imply a certain form of ownership. The term 'programme' was used instead of 'development type' to clarify that the term encompasses more than physical developments. 'Land uses' in the ontology have more general names which are zoning-type specific, such as 'Commercial use hotel zone', and hence refer to sets of programmes. Each 'land use' is related to one zoning type in Singapore, and can be linked to a unique use quantum and/or set of allowed programmes. Land uses are not generalisable or transferrable to other contexts or countries (though new 'land uses' could be defined to fill this need). 'Programmes' are universal, and can be reused in any context.

In addition to URA documents, programmes in the ontology are derived from two other sources, Google Places and the LBCSv2 ontology (see Section 3.2). The first reason for this is that, the programmes found in the URA documents range from broad to specific, infinitely divisible, and are only referred to as 'examples' in the URA documents, implying that they are non-exhaustive (see Fig. 2). Secondly, a wider range of programmes supports the goal of linking zoning types to a comprehensive list of programme types to enable future urban analyses, URA programmes were supplemented with ones from Google Places are 'establishments, geographic locations, or prominent points of interest' included in Google Maps, which are each tagged with a 'Place type' (Google, 2022). Compared to the URA documents, Google Place types provide a more comprehensive list of consumer-facing establishments occurring in Singapore, such as gyms and libraries (Google, 2022). Activities found in the LBCSv2 ontology, which represent the observable activities taking place on the land, were also added. The ontology contains particularly many classes for recreational activities, and thus it complements the programmes sourced from the URA and Google well. At this stage, even overlapping concepts were collected.

In sum, two concepts, 'land use' and 'programme', were sourced from three sources: URA documents, Google Places and LBCSv2 activities. For example, 'CommercialOrHotelUseCommercialZone' is a land use that is allowed only in Commercial zones, has a minimum use quantum of 60% of the plot's gross floor area, and allows programmes including 'Restaurant', 'Hotel', and 'Bookstore.' There is a finite number of land uses that are allowed in each zoning type, but each land use could allow an almost infinite set of programme combinations. Programmes are universal concepts that may be found in any country, while the zoning types, land uses and use quantum in OntoZoning are specific to Singapore. To collect all necessary terms and relationships to include within these superclasses and

between them, the URA documents, the list of Google Places and the LBCSv2 activities were manually scanned. This resulted in lists of zoning types, land use types, programme types and various relationships ('allows', 'may allow', etc.) found in the source data.

3.4. Creating a concept network

The next step was to organise the lists of concepts and relationships into a network format, which can be expressed as a spreadsheet or a diagram (for an excerpt, see Fig. 3). This network should represent all required knowledge.

Some terms in the lists of land uses and programmes were removed in the case of contradictions or ambiguity. For example, the Masterplan Written Statement says that the Business 1 zoning type allows the land use 'Light industry.' However, the Development Handbook for the Business 1 zoning type contain no mention of 'light industry', but list more specific types of allowed programmes such as 'core media' or 'research and development', which collectively have a use quantum attached to them (Urban Redevelopment Authority, 2019a; 2019b). It is unclear whether the use quantum also applies to 'light industry', and what the relation between 'light industry' and the more specific programmes is, so the term was left out. Similarly, programmes found in Google and LBCSv2 (such as 'Logging' in LBCSv2) which could not be reliably linked to zoning types in Singapore were left out. Ultimately, 54 programmes from Google and 72 programmes from LBCSv2 were added to the ontology. All of these were linked to land uses through the 'may allow' relationship, since based on the URA documents, it was not certain whether the land uses allowed in zoning types allow those programmes.

Organising the programmes into a hierarchy was considered. Given the large number of potential programmes, a hierarchy is useful as it facilitates finding relevant programmes. However, forming a consistent general hierarchy is difficult, or even impossible. First, it is sometimes difficult to determine what concept is more general than another. For example, should 'Recreation club' be a subtype of 'Sports and recreation facility', and should an activity like 'Tennis' be a subtype of either? Second, a programme can be a subtype of several programmes, which implies the hierarchy cannot be a strict taxonomy (i.e. a tree structure). For example, a 'Bar' can be a subprogramme of both a Golf course and a Hotel (according to the URA documents). For these reasons, no hierarchy was implemented at this point. Where possible, however, programme hierarchy was reflected in the name (e.g. a meeting room within a hotel was named 'Hotel meeting room') or in the annotation (e.g. 'Media post production and distribution' was given the annotation 'a type of Core Media programme'.)

Relationships between concepts were also modelled at this stage. Each zoning type is linked to land uses through three kinds of relationships found in URA documents: 'allows', 'may allow', and (rarely) 'does not allow' (Urban Redevelopment Authority, 2019b). However, in the URA documents, these statements naturally come with qualifiers such as 'mainly' or 'subject to the evaluation of the competent authority' (Urban Redevelopment Authority, 2019b). These qualifiers were noted down but not included in the final ontology.

After this filtering process, we linked zoning types to land use types through three relationships: 'allows use', 'may allow use', and 'does not allow use.' Land uses and programmes are similarly linked through the relationships 'allows programme' (for programmes explicitly mentioned by the URA) and 'may allow programme' (for programmes sourced from our non-URA sources, Google and the LBCSv2 ontology). Land uses are linked to use quanta through the relationships 'has minimum use quantum' and 'has maximum use quantum', and use quanta have properties expressing their value, unit, and the kind of entity they are relative to. For example, in Fig. 3, the land use shown (commercial or hotel use in commercial zone) has a minimum use quantum of 60% of the whole plot's gross floor area. Sometimes use quanta may also be absolute values (e.g. 3000 m²), or percentages that are relative to other land uses

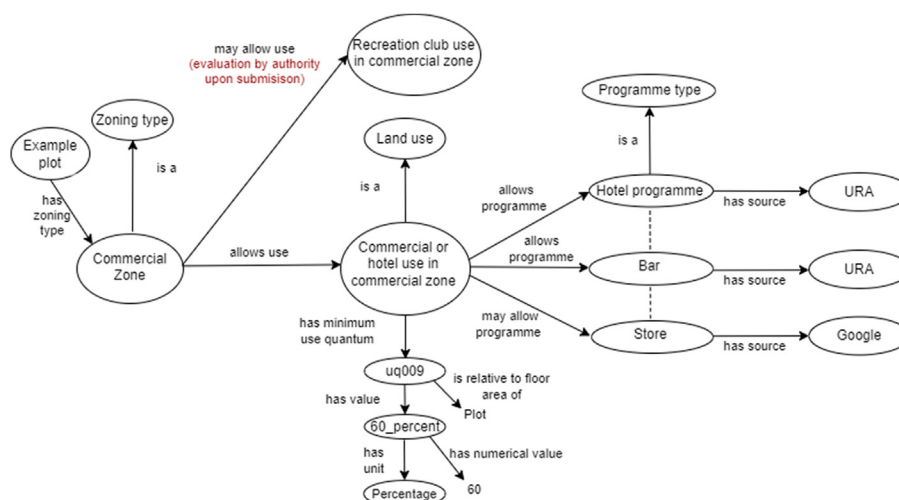


Fig. 3. An excerpt from the conceptual ontology, illustrating a selection of key concepts and relationships. The information marked in red (i.e. evaluation by authority upon submission) was not included in the final ontology. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

allowed on the plot.

Following the open world assumption, only those relationships explicitly mentioned by the URA were modelled. In other words, if the URA documents state that certain land uses are allowed in a zoning type, then it was not assumed that all other land uses are *not* allowed in that zoning type. The ‘does not allow’ relation was added only in the few cases where the URA explicitly stated that a certain land use or programme was not allowed.

3.5. Creating and evaluating the logical ontology

Next, the spreadsheets created as part of the conceptual ontology were translated into a logical ontology using the Owlready 2 Python library (Lamy, 2017) and Protégé, an open-source ontology editing software (Musen, 2015). In this step, the conceptual ontology’s concepts are translated into classes or individuals, relationships into object properties, and attributes into data properties, each with a unique identifier.

The consistency of the ontology was checked using Protégé’s Hermit reasoner. The accuracy of the ontology was checked with Protégé’s Debugger plugin. At this point, the ontology was also tested through Description logic queries. The debugger tests were derived from our competency questions 1–3, and are listed in Appendix A.3.

3.6. Linking the ontology to geospatial data

Answering the competency questions requires links to geospatial data. Besides the URA Master Plan plot data, several other datasets were used, such as building footprints. The data were converted from KML into CityGML format and uploaded into the Cities Knowledge Graph, which uses Blazegraph as its graph database (Chadzynski et al., 2021).

The zoning types, land use types and programme types in the OntoZoning ontology were then linked to the geospatial plots based on each plot’s zoning type in the raw data. Specifically, the Python RDFlib library (Boettiger, 2018) was used to translate the same spreadsheets used to generate the ontology (see Section 3.4) into an RDF file consisting of subject-predicate-object-graph quads. These quads express the concepts and relationships included in the ontology (e.g. ‘zoning type’ ‘allows use’ ‘land use’), and also the links between zoning types and the URIs of each plot in Singapore. This RDF file was then loaded into Blazegraph.

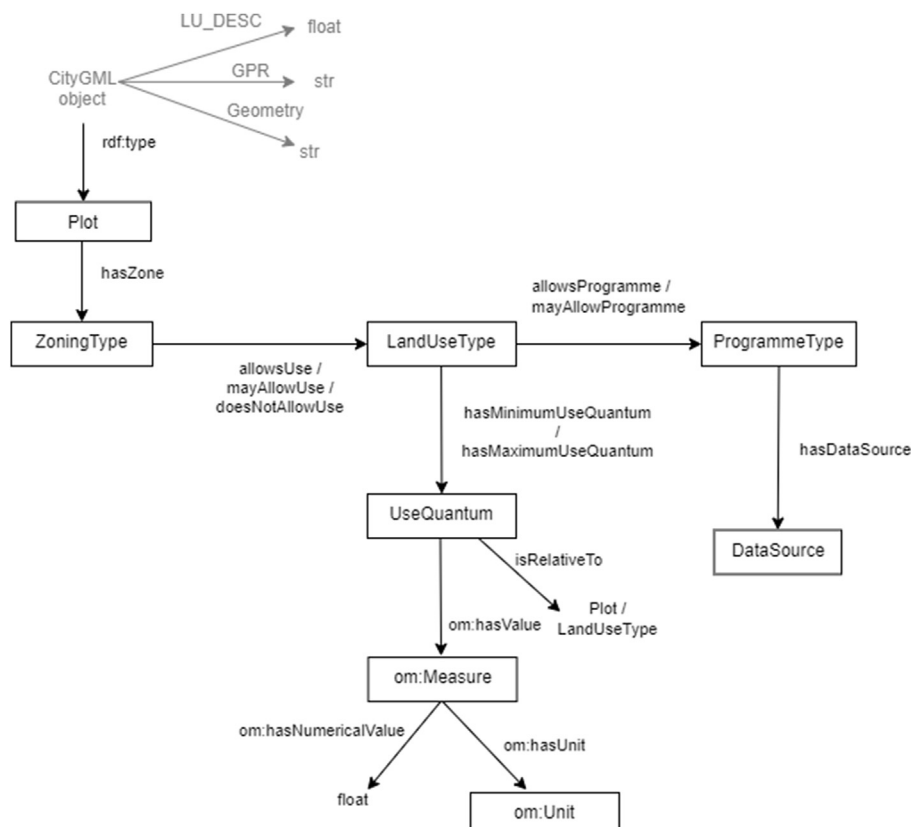


Fig. 4. Diagram illustrating the zoning and land use ontology. It is possible to link the ontology to geospatial data: CityGML geospatial objects (grey) may be classified as Plots with a certain ZoningType that is known based on the object’s LU_DESC value (short for land use description). Classes and relations with the prefix ‘om’ are imported from the Ontology of Measures (Rijgersberg et al., 2013).

A Python script was then created to query, analyse and visualise the data described above, using a SPARQL wrapper and the Geopandas and Contextily libraries (Arribas-Bel & distributors, 2020; Kelsey Jordahl et al., n.d.). While the goal is to fully integrate data analysis and visualisation into the Cities Knowledge Graph architecture, the Python scripts used to generate the present results did not interact with the knowledge graph after the query stage; rather, queries facilitated the retrieval of data for further analysis and visualisation.

4. Results and demonstrations

This section presents the zoning ontology and six proof of concept demonstrators of multi-criteria site selection. The latter show how the ontology can be used to query urban planning-related data stored in a knowledge graph, in order to answer competency questions 1–5.

4.1. The OntoZoning ontology

The ontology represents six core classes: ‘Plot’, ‘Zoning type’, ‘Land use type’, ‘Use Quantum’, ‘Programme type’ and ‘Data source’. These classes, as well as their relationships, are shown in Fig. 4. There are 32 distinct zoning types, 108 land use types, 372 programme types, 55 use quants and three data sources. In addition, two classes are imported from the Ontology of Measures: ‘Unit’ and ‘Measure’ (Rijgersberg et al., 2013).

The ontology may be linked to geospatial CityGML objects by classifying these objects as plots with a certain zoning type. The CityGML objects are derived from master plan data from the URA, and have an attribute called ‘LU_DESC’ (short for ‘land use description’) which is equivalent to the zoning type of the plot.

4.2. Query 1: uses allowed in a plot

The first query shows which uses and programmes are allowed in a plot with a certain URI (Fig. 5), addressing competency question 1. The plot in the query has zoning type ‘Sports and Recreation’, which allows seven uses: ‘Recreation club use’, ‘Marina use’, etc. Each of

```

1 PREFIX ogcml: <http://www.theworldavatar.com/ontology/ontocitygml/citieskg/OntoCityGML.owl#>
2 PREFIX zo: <http://www.theworldavatar.com/ontology/ontozoning/OntoZoning.owl#>
3 SELECT ?zone ?use ?programme
4 WHERE {
5   GRAPH <http://127.0.0.1:9999/blazegraph/namespace/singaporeEP5G4326/sparql/ontozone> {
6     <http://127.0.0.1:9999/blazegraph/namespace/singaporeEP5G4326/sparql/cityobject/UUID_9195a574-bca8-4deb-9151-a23cad79aad1>
7       zo:hasZone ?zone.
8       ?zone zo:allowsUse | zo:mayAllowUse ?use.
9       ?use zo:allowsProgramme | zo:mayAllowProgramme ?programme.
10
11     BIND (strafter(STR(?zone_uri), "#") as ?zone).
12     BIND (strafter(STR(?use_uri), "#") as ?use).
13     BIND (strafter(STR(?programme_uri), "#") as ?programme) }
14   }
15 } GROUP BY ?use ?programme ?zone
16 .

```

zone	use	programme
SportsAndRecreation	GolfUse	Restaurant
SportsAndRecreation	GolfUse	Chalet
SportsAndRecreation	GolfUse	GolfCourse
SportsAndRecreation	GolfUse	Bar
SportsAndRecreation	GolfUse	ClubHouse
SportsAndRecreation	GolfUse	DrivingRange
SportsAndRecreation	GolfUse	Golf
SportsAndRecreation	GolfUse	Guesthouse
SportsAndRecreation	GolfUse	Lounge
SportsAndRecreation	GolfUse	ProShop
SportsAndRecreation	MarinaUse	Marina
SportsAndRecreation	MarinaUse	BoatMooring,Docking,Serviceing
SportsAndRecreation	MarinaUse	Water-skiing
SportsAndRecreation	MarinaUse	Boating,Sailing,Etc
SportsAndRecreation	MarinaUse	Canoeing,Kayaking,Etc
SportsAndRecreation	MarinaUse	Fishing,Angling,Etc
SportsAndRecreation	MarinaUse	Port,Ship-building,RelatedActivities
SportsAndRecreation	MarinaUse	Sailing,Boating,OtherPort,MarineAndWater-basedActivities
SportsAndRecreation	RecreationClubUse	RecreationClub
SportsAndRecreation	RecreationClubUse	BowlingAlley

Fig. 5. A screenshot of the BlazeGraph interface which is used to query data stored in a knowledge graph through SPARQL. The top of the image shows Query 1, asking for a plot's zoning type, allowed uses, and allowed programmes per use. An excerpt of the results is shown at the bottom.

these uses allows or may allow distinct sets of programmes, such as 'BowlingAlley', from Google Maps, and 'Water-skiing', from the LBCSv2 ontology. Thus the query shows how uses (which are always derived from URA documents and which are each linked to only one zoning type) can be linked to specific programmes, which may come from other sources. This query only requires data on plot IDs, their zoning types, the uses allowed in each zoning type, and the programmes that each use contains. As such, it is the most basic demonstration of the ontology.

4.3. Query 2: plots allowing a combination of uses

Our second demonstration is a query for all plots which allow developments containing both a printing press and tennis activities, addressing competency question 2. Although this combination is allowed in multiple zoning types, it is not intuitively obvious which zoning types do so, and manually going through the allowed uses for each type would require significant effort. Querying the zoning types that allow these programmes is easier. Unlike the first query, this query uses geospatial plot data across Singapore, which makes it possible to export the query result for visualisation, as shown in Fig. 6.

4.4. Query 3: use quantum

The third query searches for information on how much of the programme 'Hotel' could be built on plots that allow both 'Hotel' and 'Test laboratory' programmes to be built. From a separate graph stored in the knowledge graph, the query first retrieves plots' buildable area (gross floor area), and filters out plots that have zero buildable area. Then, the query retrieves the relevant zoning types, the land use types that allow hotels, and the use quantum of those land uses. The query results in Fig. 7 show how the query returns all information required to answer the original question: the plot's buildable area and the use quantum information, which include a numeric value, a unit, and a related entity (the latter is used to calculate the area associated with the use quantum, if the use quantum's unit is a percentage). Based on this information, it is easy to calculate the allowed area of hotels on each plot using a script or Excel. Thus, although this query does not retrieve the allowed floor area for hotels directly, it facilitates finding this information compared to manually cross-referencing multiple relational databases.

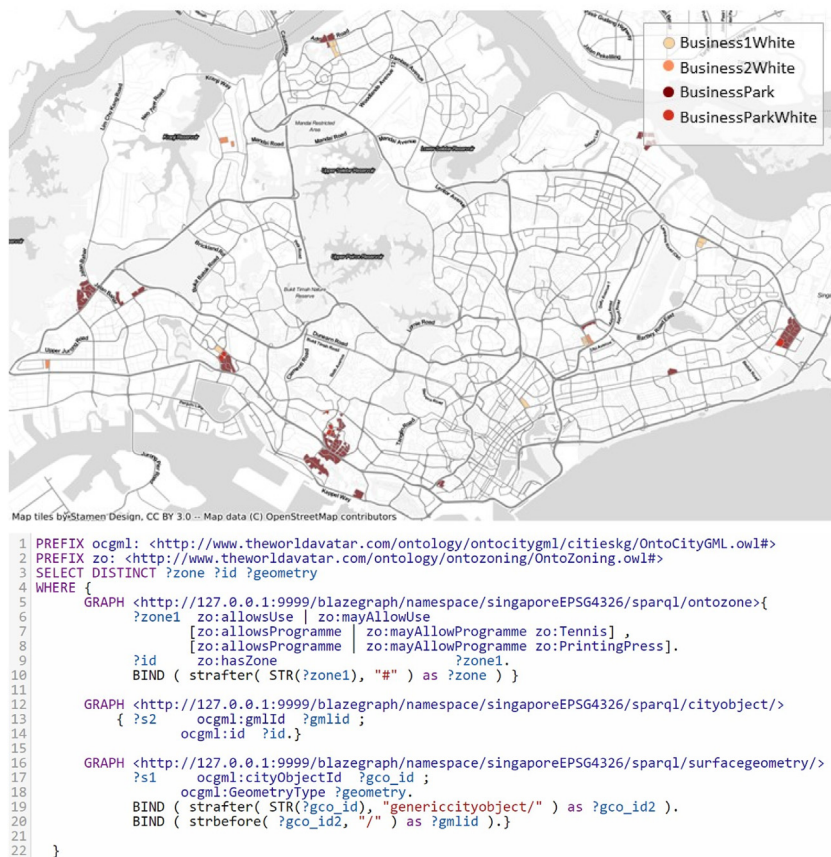


Fig. 6. Visualisation of the results of the second query (bottom), showing plots allowing a printing press and tennis activities, coloured by zoning type. The map is focused on the centre of Singapore. The query integrates OntoZoning data with plot geometry.

```

1 PREFIX ocm1: <http://localhost/ontocitygml/>
2 PREFIX oz: <http://www.theworldavatar.com/ontology/ontozoning/OntoZoning.owl#>
3 PREFIX om: <http://www.ontology-of-units-of-measure.org/resource/om-2/>
4 PREFIX pco: <http://www.theworldavatar.com/ontology/planningconceptonto/PlanningConceptOntology.owl#>
5
6 SELECT DISTINCT ?plot ?buildable_area ?land_use ?hotel_max_quantum_value ?hotel_max_quantum_unit ?hotel_max_quantum_relative_to
7 WHERE {
8   GRAPH <http://www.theworldavatar.com:83/citieskg/namespace/singaporeEPSG4326/sparql/buildablespace/> {
9     ?plot uri
10    ?buildable_space pco:hasAllowedGFA ?buildable_space.
11    ?gfa om:hasValue ?gfa_measure.
12    ?gfa_measure om:hasNumericValue ?buildable_area.
13
14    BIND (strafter( STR(?plot_uri), 'cityobject/') as ?plot).}
15
16   GRAPH <http://www.theworldavatar.com:83/citieskg/namespace/singaporeEPSG4326/sparql/ontozoning/> {
17     ?zone_uri oz:allowsUse [oz:allowsProgramme oz:TestLaboratory], [oz:allowsProgramme oz:HotelProgramme].
18     ?zone_uri oz:allowsUse ?land_use_uri.
19     ?land_use_uri oz:allowsProgramme oz:HotelProgramme.
20     ?land_use_uri oz:hasMaxUseQuantum ?uq.
21     ?uq oz:isRelativeTo ?hotel_max_quantum_relative_to.
22     ?uq om:hasValue ?uq_measure.
23     ?uq_measure om:hasNumericalValue ?hotel_max_quantum_value.
24     ?uq_measure om:hasUnit ?hotel_max_quantum_unit_uri.
25
26     BIND (strafter( STR(?zone_uri), '#') as ?zoning_type).
27     BIND (strafter( STR(?land_use_uri), '#') as ?land_use).
28     BIND (strafter( STR(?hotel_max_quantum_unit_uri), 'om-2/') as ?hotel_max_quantum_unit).}
29 }

```

plot	buildable_area	land_use	hotel_max_quantum_value	hotel_max_quantum_unit	hotel_max_quantum_relative_to
UUID_cd37edc0-a8c-423e-b98c-bc009db79dc6/	14049.498992193105	WhiteUseBusinessParkWhite	40.0	Percentage	Plot
UUID_cd37edc0-a8c-423e-b98c-bc009db79dc6/	14049.498992193105	WhiteUseBusinessPark	15.0	Percentage	Plot
UUID_9183c22c-0c39-4cf2-85bf-901bf9f9b996/	2919.982744469076	WhiteUseBusinessParkWhite	40.0	Percentage	Plot

Fig. 7. Screenshot of Query 3 and its results. The query retrieves information on how much buildable area can be allocated to hotels, on plots that allow both a hotel and a showroom. Based on the plot's overall buildable area and the use quantum information (in this case, expressed as a percentage of the plot's buildable area), the allowable hotel area can be easily calculated.

4.5. Query 4: adding geospatial conditions

The fourth query searches for plots that 1) are located within 1000 m of an MRT (mass rapid transit) station entrance; 2) contain no existing buildings; and 3) allow or may allow building a gym and a hotel. The result is visualised in Fig. 8. It addresses competency question 2: 'where can I place a specific programme, or mix of programmes?' This example also demonstrates how further geospatial conditions (proximity to MRT), stemming from different urban datasets, can be incorporated in a query. The results of these topological analyses and filters were not obtained directly from the knowledge graph, but rather knowledge graph queries were used to retrieve the information to simplify processing.

4.6. Query 5: adding geospatial conditions and new datasets

This is a query to find plots that allow residential development and have unused gross floor area (GFA), and are located adjacent to a park. Unused GFA was approximated for each plot by multiplying the area of each building footprint found on the plot with the height of the building, and subtracting the result from the total GFA allowed on the plot. Building height data originates from a dataset described in Disegna et al. (2019), which was stored in the knowledge graph. Total GFA was calculated by multiplying the plot's area and GPR, as

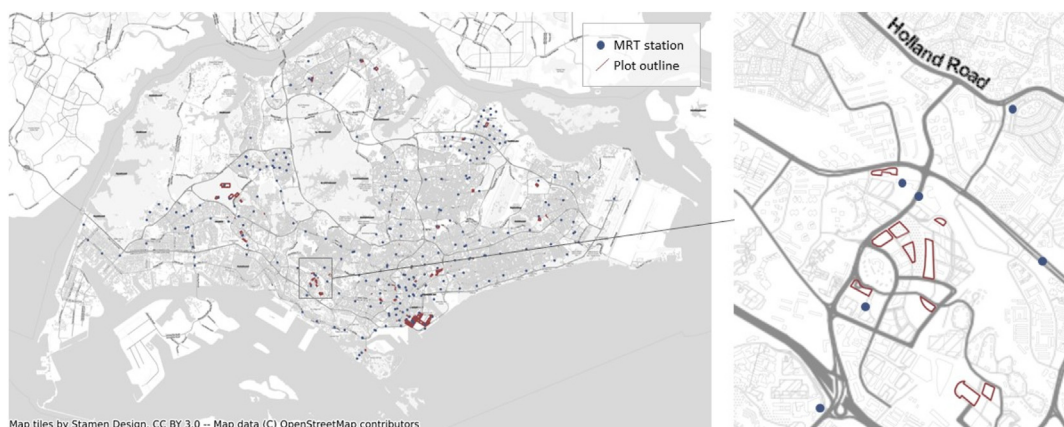


Fig. 8. Visualisation of Query 3, showing empty plots located within 1000 m of an MRT station, that allow building a gym and hotel development.

specified in the Master Plan. The query result is visualised in Fig. 9. This query also addresses competency question 2. As before, topology analyses were conducted separately based on data downloaded from the knowledge graph.

4.7. Conceptualisation of links to further regulatory data

Answering competency question 5 in the future requires considering how OntoZoning could ultimately be integrated within the broader regulatory system, in which land use-related regulations may affect plots based on other properties besides zoning type. Fig. 10 explores the potential for this integration through a conceptual diagram that extends the Query 3 example related to use quantum. The diagram represents the link between OntoZoning and a regulation according to which the use quantum for 'White' uses can be higher for plots zoned 'Business Park' if the plot is located next to an MRT station (Urban Redevelopment Authority, 2019d). In this case, data about the use quantum exception could be added to the knowledge graph and linked to all relevant plots that have as their 'has location' and 'has zoning type' property values 'next to MRT' and 'Business Park.' Then, when searching for the plots' use quantum, one could also query for use quantum exceptions that apply to the plot. Other land use planning regulations could be matched to plots similarly.

5. Discussion

What opportunities and challenges do semantic web approaches present for multi-criteria site selection, and urban regulatory data more broadly? In terms of opportunities, linking land use regulations to geospatial data through ontologies facilitates access to relevant land use regulatory information, as demonstrated by the multi-criteria site selection queries. These data access and usability benefits would increase as further regulatory and other data is added to the knowledge graph. The present work suggests that it is feasible to represent a key part of a country's land use regulations using ontologies, including more intricate exceptions-based regulations. However, the semantic approach may also introduce challenges related to the management and maintenance of urban planning regulatory systems.

Semantic web approaches promise benefits compared to the status quo in terms of accessing and using land use regulation data. Currently, urban planning documents, including those used in Singapore, come from several sources, have different formats, and contain overlapping information. Literature on planning regulations suggests this is a common issue globally (Chen et al., 2018; Dong et al., 2017; Hirt, 2012; Thorsby et al., 2017). Representing regulatory data in ontologies and storing them in knowledge graphs allows accessing the data through SPARQL queries, which are arguably easier and faster than searches on the original websites, datasets and documents.

Section 4 of this paper also suggested that SPARQL queries on a knowledge graph may be easier and faster in comparison to queries or scripts operating on traditional relational databases. This is because the semantic web approach allows querying land use-related and geospatial data from a single end point, using the simple triple data structure. There is no need to consider the structure of the data, and all efforts to build and use interoperable datasets can focus on ensuring semantic compatibility (Kuhn et al., 2014). In contrast, in a relational database, the land use data represented in the ontology would need to be represented in several different tables, and the plot data would be distributed among several other datasets. Writing scripts to link these data, or integrating these datasets in GIS, would require not only understanding of the semantic links between the concepts, but also the database structures.

This benefit of interoperability in the semantic web approach becomes more apparent the more complex the data structures are, the greater their number, and the more comprehensively they represent the domain of interest. Although the OntoZoning ontology is still relatively non-complex, its use quantum relationships perhaps best illustrate the benefit of queries for easy retrieval of complex data. Query 3 demonstrated how plots' buildable space data can be combined with the OntoZoning ontology's zoning type, land use and use quantum data. Although it was not yet possible to calculate allowed floor area per plot using only data on the knowledge graph, even making a query that brings all the necessary calculation parameters together would greatly facilitate calculating the floor areas per allowed uses, compared to retrieving this information from several disparate relational datasets.

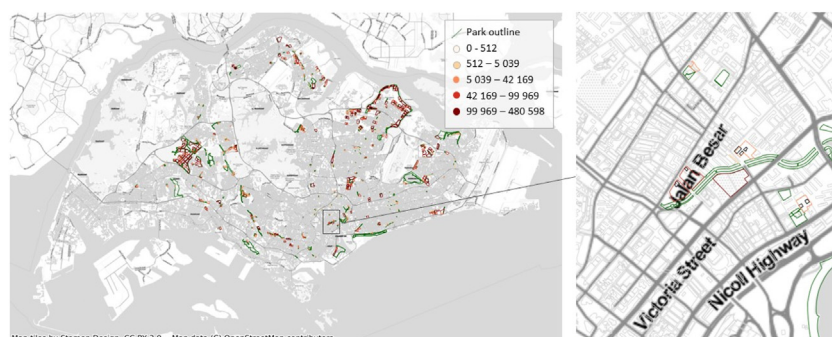


Fig. 9. Visualisation of Query 4, showing plots that allow residential development, are adjacent to a park, and have unused GFA. Parks adjacent to the plots of interest are shown in green, while plots are coloured according to the amount of unused GFA (m^2), shown in the legend. Buildings located within the selected plots have black outlines, while other buildings' outlines are grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

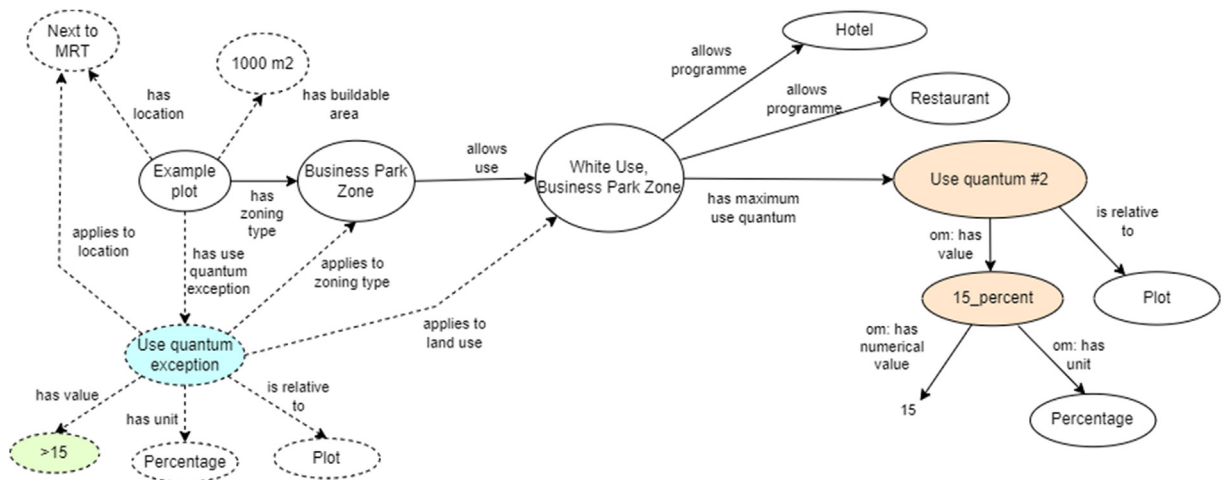


Fig. 10. Conceptual diagram showing how location-specific regulations that affect land uses could be integrated with OntoZoning. Here, a use quantum exception (turquoise) applies to the plot, and the zoning type's use quantum (orange) must be replaced with the exception (green). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Queries 4–5 hint at further questions that could be answered if OntoZoning were integrated with other urban data, such as topological relations, building footprints and the locations of mass rapid transit stations. The analyses and visualisations of Queries 4–5 were created using Python, based on data retrieved from the knowledge graph. As with Query 3 mentioned above, the ontology facilitated the access to and use of relevant urban data, even if the final results and visualisations were not obtained directly from the knowledge graph.

While these examples show that data structured in ontologies and stored in knowledge graphs is relatively easy to access, the question remains whether this benefit outweighs the effort involved in creating and maintaining such ontologies and data structures. Creating the OntoZoning ontology required a labour-intensive process of collecting terms from multiple documents, defining them, and conceptualising their relations in a way that tries to ensure compatibility with regulations that were not yet represented in the ontology. A corresponding workflow for creating and accessing data in a relational database was not tested. Likewise, the process of updating the ontology was not explored. Will the ontology be flexible enough to accommodate changes to regulations that inevitably arise in the real world? Will planners or external ontology and knowledge graph experts be in charge of maintaining the system? This question raises further questions about familiar risks related to digitisation of governance: path-dependency, vendor lock-in, and the loss of control over governance to private entities or algorithms (Brauneis and Goodman, 2018).

Based on the present work, it is not possible to conclusively weigh the costs and benefits of the semantic web approach. Evaluating whether the ontology is feasible to maintain in practice is difficult without in-depth knowledge of the resources and workflows of the URA or other urban planning agencies. The associated risks of lock-in and the prospect of loss of control (or loss of understanding) also depend on the implementation of the complete regulation ontology, the types of updates made, and whether or not the ontologies are developed in close consultation with planners or through an artificial intelligence-based process that is dependent on potentially untransparent algorithms.

6. Conclusion

Semantic web technologies may improve access to and usability of urban data, which are constantly proliferating. As such, they could contribute to a paradigm shift in planning characterised by participation, digitisation, intelligence and interconnectedness (Potts, 2020). At present, awareness of this potential in the field appears to be low, as shown by the relative lack of literature at the intersection of planning regulations and semantic web technologies. This article has sought to demonstrate on a concrete level how ontologies and semantic web technologies can benefit stakeholders of urban planning regulatory data such as developers, planners, researchers and citizens. In addition, the OntoZoning ontology developed as part of this work was designed to be reused and extended by others who wish to represent urban data semantically.

The present work has four main limitations. First, the study is specific to Singapore. This limitation is mitigated by the universality of the 'programme' classes, the relative similarity of Singapore's zoning system to those of other countries (as reviewed by Hirt, 2012) and the flexibility of the ontology structure, which consists of zoning types, land use types and programmes. Together, these features facilitate the application of OntoZoning to other countries. In addition, we have discussed how the broader problem statement of inaccessible and difficult-to-use regulatory data is applicable to other contexts (Chen et al., 2018; Dong et al., 2017; Hirt, 2012; Thorsby et al., 2017). A second limitation is the relative lack of user-friendliness in the ontology and the workflow presented in this paper. The ontology contains many distinct programmes, and the concept of 'land use' as a set of programmes that is specific to a zoning type may seem unintuitive to those unfamiliar with the ontology. When querying the ontology, automatic visualisation of the results would also facilitate the use of the ontology. Thirdly, OntoZoning only represents the core part of land use planning regulations of Singapore. It excludes certain regulations that affect allowed land uses on a plot, depending on its context in e.g. a certain area. Finally, the fourth

limitation is that the ontology has not yet been validated by URA domain experts. Thus OntoZoning is best used as a starting point to explore allowed programmes in a plot or zoning type, and not as a full and accurate reflection of Singapore's land-use regulation system.

Ongoing and future work by the Cities Knowledge Graph team, seek to address these limitations, firstly by studying the applicability of the approach to other contexts, and secondly by developing a user interface that allows querying and visualising the contents of the knowledge graph from a single endpoint (Chadzynski et al., 2022b). A third goal is to more comprehensively and accurately represent Singapore's urban planning regulations, including e.g. height control and area-specific regulations, and add the capacity to evaluate their impacts. Finally, it is also important to seek input from the urban planners and regulators, and to consider the practical implications of semantic web technologies for their workflow and required output.

Data and codes availability statement

The data and codes that support the findings of this study are available on Figshare with the identifiers <https://doi.org/10.6084/m9.figshare.21516207.v1> (ontology owl file) and <https://doi.org/10.6084/m9.figshare.21516561.v1> (scripts and data).

Declaration of competing interest

The authors report that there are no competing interests to declare.

Acknowledgements

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

Appendix

A.1 Background on methods for creating ontologies

Given that our ontology seeks to integrate data from different Singaporean zoning documents, it is most feasible to construct it manually, based on domain expertise. This approach contrasts with various automated or semi-automated methods based on clustering and natural language processing techniques, which are typically used when the input dataset is very large (Bedini, 2007). In the manual method, the concepts and relationships in the ontology are modelled by a human based on their understanding of a domain. This approach consists of two steps: building a 'conceptual ontology' understandable to humans, and translating it into a final, logical ontology interpretable by machines using description logics serialised as OWL (Mizen et al., 2005). This separation of concerns is useful for two reasons: it allows 1) keeping track of all relevant concepts and relationships, which cannot always be represented in the final ontology due to limitations in the expressivity of description logics (Mizen et al., 2005); and 2) ensuring that the final ontology is logically correct, even if the domain contains relationships which could lead to logical errors.

The conceptual ontology is created by first defining the purpose and requirements of the ontology, then collecting necessary information to be encoded in the ontology, and finally organising that information into networks of concepts, relationships and attributes (Mizen et al., 2005). In this network, concepts are typically based on nouns found in the source information, while relationships are based on verbs linking two concepts, and attributes are based on verbs or adjectives related to only one concept (examples of attributes could include having a height or an age, if these are not included as separate concepts in the ontology).

After this network of concepts is validated against the purpose of the conceptual ontology, it can be translated into a logical ontology using description logics. The logical ontology is typically built using specialised software such as Protégé, where concepts are translated into classes and individuals, relationships into object properties, and attributes into data properties, each with a unique identifier.

Since creating ontologies is labour-intensive, reusing existing ontologies is always recommended. Aside from saving resources, reuse supports the development of shared languages and interoperability described in the previous section (Katsumi & Fox, 2019).

A.2 Existing land use and planning ontologies that were not reused

(1) Land use ontology for St Petersburg (Chichkova et al., 2020)

- Purpose and contents. This ontology aims to answer two types of questions in the context of St Petersburg, Russia: 1) what kinds of restrictions on construction exist on a plot, and 2) which plots allow constructing a specific development, such as a factory. The ontology contains territorial zones, zoning types, permission types, land use types, and classes related to construction regulation, such as maximum number of floors. The purpose of this ontology is most similar to ours.
- Reusability. The zoning types are specific to the St Petersburg context, and hence not reusable. The ontology's 156 land use types are in Russian, making them difficult to reuse. The PermissionTypes ('primarily permitted', 'conditionally permitted', 'auxiliary permitted') are very few in number, limiting the benefits of reuse.

(2) Land use ontology for Morocco (Barramou et al., 2020)

- Purpose and contents. This ontology represents land uses allowed in Morocco's zoning types. It contains classes belonging to 4 categories: zoning, services, infrastructure and easement.

- Reusability. Of these categories, zoning types cannot be reused because they are different in Morocco compared to Singapore. Infrastructure and easement are not within the scope of our ontology, so there is no need to reuse these classes. The ten services in this ontology are more general (including ‘Health’, ‘Education’, ‘Parking’, ‘Commercial’, etc.). Although these are applicable to the Singapore context, we found that they are too broad and few in number for our purposes. Allowing stakeholders to search for a plot for a specific use or mix of uses, or the kinds of transportation and energy analyses envisioned in the Introduction, requires a more detailed list of services. Thus, we opted not to reuse this ontology.
- (3) Land use ontology for Taiwan (Kuo & Hong, 2012)
- Purpose and contents. Allowing users to retrieve land use data and semantic data linked to City GML data. The ontology contains land use classes such as ‘Agriculture’, ‘Public use’, or ‘Traffic’, as well as more specific subclasses for each one.
 - Reusability. Most of the classes in this ontology are not relevant to an urban development context, as the focus is more on agricultural land uses. In addition, the ontology is not available online, to the best of our knowledge. For these reasons, this ontology was not reused.
- (4) OSMOnto (Codescu et al., 2011)
- Purpose and contents. The purpose of this ontology is to link OSM tags for map locations to an ontology of activities. Rather than representing a domain, the aim is to facilitate the task of querying where to find activities of interest on the map, for example when planning a route. The ontology thus links OSM tags and activities. The Activity ontology (which is more relevant for our purposes) contains a wide range of classes, e.g. ‘Charging station’, ‘Civil service’ and ‘Restaurant’.
 - Reusability. Despite this ontology’s high granularity of activities, we opted not to reuse it. This is mainly because the ontology has a very different purpose than ours, which means that many of its activity classes and properties are irrelevant from our perspective. For example, the class ‘Restaurant’ has the property ‘hasCuisineOfNationality’, linking it to different countries. Another example of an overly granulated class is the Medicine class, which is divided into a large number of specialisations (including nine subfields of internal medicine, among others).

A.3 OntoZoning debugger queries

The ontology was tested using the Protege debugger and through description logic queries (in the DL queries tab in Protégé). Table 1 below contains a list of entailed test cases for the Protégé Debugger, organised by category. Table 2 contains a list of non-entailed test-cases.

Table 1
Entailed test cases for the Protégé debugger.

Zoning types and land uses		
White	allowsUse	PrivateMedicalClinicWhiteZone
Residential	allowsUse	ServicedApartmentUseResidentialZone
BusinessPark	allowsUse	MainUseBusinessPark
CommercialOrInstitution	doesNotAllowUse	FlatOrCondominiumUseCommercialOrInstitutionZone
Land uses and programmes		
HotelUseHotelZone	allowsProgramme	HotelStaffCanteen
ServicedApartmentUseWhiteZone	allowsProgramme	ServicedApartmentMixedUseZone
ServicedApartmentUseResidentialZone	allowsProgramme	ServicedApartmentResidentialZone
CommercialUseHotelZone	mayAllowProgramme	Pharmacy
LandedHousingUse	allowsProgramme	TerraceType1
MainUseBusinessPark	allowsProgramme	VideoEditing
Use quantums		
PrivateMedicalClinicWhiteZone	hasMaxUseQuantum	uq015
uq015	hasValue	20_percent
20_percent	hasUnit	Percentage
20_percent	hasNumericalValue	20
uq015	isRelativeTo	“CommercialUseWhiteZone”
uq012	isRelativeTo	“Plot”
Programmes and sources		
Bank	hasSource	URA
Bank	hasSource	Google
SeismicDataAnalysisCentre	hasSource	URA
MediaLibraryServices	hasSource	URA

Table 2
Examples of non-entailed test cases for the Protégé debugger.

Zoning types and land uses		
White	doesNotAllowUse	FlatOrCondominiumUseCommercial OrInstitutionZone
Land uses and programmes		
HotelUseHotelZone	allowsProgramme	VideoEditing
Use quantums		
PrivateMedicalClinicWhiteZone	hasMaxUseQuantum	uq40
20_percent	hasNumericalValue	40
Programmes and sources		
Clinic	hasSource	Google

References

- Ahlqvist, O., Varanka, D., & Fritz, X. (2016). *Land use and land cover semantics: Principles, best practices and prospects*. In K. Janowicz (Ed.). Boca Raton: CRC Press.
- Arribas-Bel, D., & distributors, C. (2020). *Context*.
- Barramou, F., Mansouri, K., & Addou, M. (2020). Toward a multi-dimensional ontology model for urban planning, 06 *Journal of Geographic Information System*, 12, 697–715.
- Bedini, I. (2007). *Automatic Ontology Generation: State of the Art*. Retrieved September 21, 2021, from <https://www.semanticscholar.org/paper/Automatic-Ontology-Generation-%3A-State-of-the-Art-Bedini/81989605fba59a415a28603ce709566cc6ebc45c>.
- Bibby, P., & Shepherd, J. (2000). GIS, land use, and representation. *Environment and planning B: Planning and Design*, 27(4), 583–598. <https://doi.org/10.1068/b2647>
- Boettiger, C. (2018). *Rdflib: A high level wrapper around the redland package for common rdf applications*. <https://doi.org/10.5281/zenodo.1098478>.
- Brauneis, R., & Goodman, E. P. (2018). Algorithmic transparency for the smart city. *Yale Journal of Law & Technology*, 20(1), 103.
- Chadzynski, A., Krdzavac, N., Farazi, F., Lim, M. Q., Li, S., Grisiute, A., ... Kraft, M. (2021). Semantic 3D City Database—An enabler for a dynamic geospatial knowledge graph. *Energy and AI*, 6, 100106.
- Chadzynski, A., Li, S., Grisiute, A., Chua, J., Yan, J., Tai, H., ... Kraft, M. (2022b). Semantic 3D City Interfaces—intelligent interactions on Dynamic Geospatial Knowledge Graphs. *c4e-Preprint Series*, 297. <https://como.ceb.cam.ac.uk/preprints/297/>.
- Chadzynski, A., Li, S., Grisiute, A., Farazi, F., Lindberg, C., Mosbach, S., ... Kraft, M. (2022a). Semantic 3D City Agents—An intelligent automation for dynamic geospatial knowledge graphs. *Energy and AI*, 8, 100137.
- Chen, Y., Sabri, S., Rajabifard, A., & Agunbiade, M. E. (2018). An ontology-based spatial data harmonisation for urban analytics. *Computers, Environment and Urban Systems*, 72, 177–190. <https://doi.org/10.1016/j.compenvurbsys.2018.06.009>
- Chichkova, N., Begler, A., & Vlasov, V. (2020). Modeling city land use with an ontology. *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance*, 851–854.
- Codescu, M., Horsinka, G., Kutz, O., Mossakowski, T., & Rau, R. (2011). *DO-ROAM: Activity-Oriented Search and Navigation with OpenStreetMap*. Claramunt.
- Dickinson, G., & Shaw, M. (1977). What is “Land use”? *Area*, 9(1), 38–42. <https://www.jstor.org/stable/20001161>.
- Dissegna, M. A., Yin, T., Wei, S., Richards, D., & Grêt-Regamey, A. (2019). 3-D reconstruction of an urban landscape to assess the influence of vegetation in the radiative budget. *Forests*, 10(8), 700.
- Dong, H., Singh, G., Attri, A., & El Saddik, A. (2017). Open data-set of seven Canadian cities. *IEEE Access*, 5, 529–543.
- Engin, Z., van Dijk, J., Lan, T., Longley, P. A., Treleaven, P., Batty, M., & Penn, A. (2020). Data-driven urban management: Mapping the landscape. *Journal of Urban Management*, 9(2), 140–150.
- Google. (2022). *Place types - places API documentation*. Retrieved April 4, 2022, from https://developers.google.com/maps/documentation/places/web-service/supported_types.
- Gribb, W., & Czerniak, R. (2016). Land use/land cover classification systems and their relationship to land planning. In O. Ahlqvist, D. Varanka, X. Fritz, & K. Janowicz (Eds.), *Land use and land cover semantics: Principles, best practices and prospects*. Boca Raton: CRC Press.
- Grisiute, A., Silvennoinen, H., Li, S., & Chadzynski, A. (2022b). Unlocking Urban Simulation Data with a Semantic City Planning System. *Co-creating the Future: Inclusion in and through Design, Proceedings of the 40th eCAADe conference*.
- Grisiute, A., Shi, Z., Chadzynski, A., Silvennoinen, H., Richthofen, A. V., & Herthogs, P. (2022a). *Automated Semantic SWOT Analysis for City Planning Targets: Data-driven Solar Energy Potential Evaluations for Building Plots in Singapore, Proceedings of the 27th International Conference on Computer-Aided Architectural Design Research in Asia*, 1, 555–565.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220.
- Guarino, N., Oberle, D., & Staab, S. (2009). What is an ontology? In S. Staab, & R. Studer (Eds.), *Handbook on ontologies* (pp. 1–17). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-92673-3_0.
- Heath, T. (2010). A taskonomy for the semantic web. *Semantic Web*, 1(1–2), 75–81.
- Hirt, A. (2012). *Mixed use by default: How the Europeans (don't) zone*. *Journal of planning literature*, 27(375). <https://doi.org/10.1177/0885412212451029>
- Hoekstra, R., Winkels, R., & Hupkes, E. (2010). Spatial planning on the semantic web. *Transactions in GIS*, 14(2), 147–161.
- Höffner, K., Walter, S., Marx, E., Usbeck, R., Lehmann, J., Ngomo, N., & C. A. (2017). Survey on challenges of question answering in the semantic web. *Semantic Web*, 8(6), 895–920.
- Huang, Z. (2021, September 1). AI in urban planning: 3 ways it will strengthen how we plan for the future. Retrieved October 28, 2022, from <https://www.ura.gov.sg/Corporate/Resources/Ideas-and-Trends/AI-in-Urban-Planning>.
- Iwaniak, A., et al. (2016). Semantic metadata for heterogeneous spatial planning documents. *ISPRS annals of the photogrammetry. Remote Sensing and Spatial Information Sciences*, 4, 27.
- Jacob, E. K. (2003). Ontologies and the semantic web. *Bulletin of the American Society for Information Science and Technology*, 29(4), 19–19.
- K. Jordahl, Van den bossche, J., fleischmann, M., McBride, J., gerard, J., matthew perry, garcia badaracco, A., farmer, arne hjelte, & snow. (n.d.). *Geopandas/geopandas: V0.8.1 (Version v0.8.1)*.
- Kandt, J., & Batty, M. (2021). Smart cities, big data and urban policy: Towards urban analytics for the long run. *Cities*, 109, 102992.
- Katsumi, M., & Fox, M. (2018). Ontologies for transportation research: A survey. *Transportation Research Part C: Emerging Technologies*, 89, 53–82.
- Katsumi, M., & Fox, M. (2019). An ontology-based standard for transportation planning. *JOWO*.
- Katsumi, M., & Fox, M. (2020). *iCity Transportation Planning Suite of Ontologies*. University of Toronto. https://enterpriseintegrationlab.github.io/icity/iCityOntologyReport_1.2.pdf.

- Kuhn, W., Kauppinen, T., & Janowicz, K. (2014). *Linked Data - A Paradigm Shift for Geographic Information Science*. 10.1007/978-3-319-11593-1_12.
- Kuo, C.-L., & Hong, J.-H. (2012). Hierarchical ontology development and semantics retrieval for land use data. *2012 IEEE International Geoscience and Remote Sensing Symposium*, 6051–6054.
- Lamy, J.-B. (2017). Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. *Artificial Intelligence. Medicine*, 80, 11–28. <https://doi.org/10.1016/j.artmed.2017.07.002>.
- Levine, J., Grengs, J., & Merlin, L. A. (2019). *From mobility to accessibility: Transforming urban transportation and land-use planning*. Cornell University Press. <http://www.jstor.org/stable/10.7591/j.ctvfc52mj>.
- Mai, G., Janowicz, K., Cai, L., Zhu, R., Regalia, B., Yan, B., Shi, M., & Lao, N. (2020). SE-KGE: A location-aware knowledge graph embedding model for geographic question answering and spatial semantic lifting. *Transactions in GIS*, 24(3). <https://doi.org/10.1111/tgis.12629>
- Marwedel, J. (1998). Opting for performance: An alternative to conventional zoning for land use regulation. *Journal of Planning Literature*, 13(2), 220–231. <https://doi.org/10.1177/088541229801300205>
- McGuinness, D. L., & Van Harmelen, F. (2004). OWL web ontology language overview. *W3C recommendation*, 10(10), 2004.
- Mizen, H., Dolbear, C., & Hart, G. (2005). Ontology ontogeny: Understanding how an ontology is created and developed. In M. A. Rodríguez, I. Cruz, S. Levashkin, & M. J. Egenhofer (Eds.), *GeoSpatial semantics* (pp. 15–29). Berlin Heidelberg: Springer. https://doi.org/10.1007/11586180_2.
- Mohamed, A., & Worku, H. (2019). Quantification of the land use/land cover dynamics and the degree of urban growth goodness for sustainable urban land use planning in Addis Ababa and the surrounding Oromia special zone. *Journal of Urban Management*, 8(1), 145–158.
- Montenegro, N. C., Gomes, J. C., Urbano, P., & Duarte, J. P. (2012). A land use planning ontology: Lbcs. *Future Internet*, 4(1), 65–82.
- Musen, M. A. (2015). The protégé project: A look back and a look forward. *AI Matters*, 1(4), 4–12.
- Patel, A., & Jain, S. (2019). Present and future of semantic web technologies: A research statement. *International Journal of Computers and Applications*, 43(5), 413–422. <https://doi.org/10.1080/1206212X.2019.1570666>
- Portugali, J., Meyer, H., Stolk, E., & Tan, E. (2012). *Complexity theories of cities have come of age: An overview with implications to urban planning and design*. Berlin: Springer Berlin Heidelberg.
- Potts, R. (2020). Is a new ‘planning 3.0’ paradigm emerging? Exploring the relationship between digital technologies and planning theory and practice. *Planning Theory & Practice*, 21(2), 272–289. <https://doi.org/10.1080/14649357.2020.1748699>
- Rijgersberg, H., Van Assem, M., & Top, J. (2013). Ontology of units of measure and related concepts. *Semantic Web*, 4(1), 3–13.
- Rudman, R., & Bruwer, R. (2016). Defining web 3.0: Opportunities and challenges. *The Electronic Library*, 34(1), 132–154. <https://doi.org/10.1108/EL-08-2014-0140>
- Shi, Z., Silvennoinen, H., Chadzynski, A., von Richthofen, A., Kraft, M., Cairns, S., & Herthogs, P. (2022). Defining archetypes of mixed-use developments using Google Maps API data. *Environment and Planning B: Urban Analytics and City Science*, 23998083221141428.
- Soon, K. H., & Khoo, V. H. S. (2017). CityGML modelling for Singapore 3D national mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W7, 37–42.
- Stewart, T. J., Janssen, R., & van Herwijnen, M. (2004). A genetic algorithm approach to multiobjective land use planning. *Computers & Operations Research*, 31(14), 2293–2313. [https://doi.org/10.1016/S0305-0548\(03\)00188-6](https://doi.org/10.1016/S0305-0548(03)00188-6)
- Tan, Y., Wang, Z., & Zhang, Q. (2020). Land-use regulation and the intensive margin of housing supply. *Journal of Urban Economics*, 115, Article 103199. <https://doi.org/10.1016/j.jue.2019.103199>
- Thorsby, J., Stowers, G. N., Wolslegel, K., & Tumbuan, E. (2017). Understanding the content and features of open data portals in American cities. *Government Information Quarterly*, 34(1), 53–61.
- Trojahn, C., Vieira, R., Schmidt, D., Pease, A., & Guizzardi, G. (2022). ‘Foundational ontologies meet ontology matching: A survey.’. *Semantic Web*, 13(4), 685–704, 2022.
- Urban Redevelopment Authority. (2019a). *Development control: Non-residential handbooks*. Retrieved October 28, 2022, from <https://www.uragov.sg/Corporate/Guidelines/Development-Control/Non-Residential>.
- Urban Redevelopment Authority. (2019b). *Singapore Master Plan 2019*. Retrieved October 28, 2022, from <https://www.uragov.sg/Corporate/Planning/Master-Plan>.
- Urban Redevelopment Authority. (2019c). *Business Park - Use quantum*. Retrieved October 28, 2022, from <https://www.uragov.sg/Corporate/Guidelines/Development-Control/Non-Residential/Business-Park/Use-Quantum>.
- Urban Redevelopment Authority. (2019d). *Master Plan Written Statement 2019*. Retrieved October 28, 2022, from <https://www.uragov.sg/-/media/Corporate/Planning/Master-Plan/MP19writtenstatement.pdf>.
- Urban Redevelopment Authority. (2020). *Urban planning 2.0: From urban planning analytics to AI city planning assistant*.
- von Richthofen, A., Herthogs, P., Kraft, M., & Cairns, S. (2022). Semantic city planning systems (SCPS): A literature review. *Journal of Planning Literature*, 37(3), 415–432.
- Walczak, M. (2021). A multi-dimensional spatial policy model for large-scale multi-municipal Swiss contexts. *Environment and planning B: Urban Analytics and City Science*, 48(9), 2675–2690.
- Williams, J., Fox, M., & Easterbrook, S. (2021). A constraint-based decision support system for capacity planning in interdependent evolving urban systems. In *The City and Complexity Life, Design and Commerce in the Built Environment*, AMPS Proceedings Series 19.1, 301–313. AMPS.
- Winkels, R., Boer, A., & Hupkes, E. (2007). Legal atlas: Access to legal sources through maps. In *Proceedings of the 11th international conference on Artificial intelligence and law* (pp. 27–36).
- Xue, F., Wu, L., & Lu, W. (2021). Semantic enrichment of building and city information models: A ten-year review. *Advanced Engineering Informatics*, 47, 101245.
- Yazdanie, M., & Orehoung, K. (2021). Advancing urban energy system planning and modeling approaches: Gaps and solutions in perspective. *Renewable and Sustainable Energy Reviews*, 137, Article 110607.
- Zhang, Y., & Schnabel, M. A. (2017). Workflow of data integrating and parametric modelling in urban design regulation. *Back to the future: The next 50 years (51st international conference of the architectural science association (ANZAScA))*, 139–148.