Title: Infrastructure and Cities Ontologies

Authors: Liz Varga, Lauren McMillan, Stephen Hallett, Tom Russell, Luke Smith, Ian Truckell, Andrey Postnikov, Sunil Rodger, Noel Vizcaino, Bethan Perkins, Brian Matthews, Nik Lomax

Keywords: City-scale Infrastructure Operations, City-scale Simulations and Data Analytics, Urban Infrastructure Development, Critical Infrastructure, Data Analytics for Infrastructure

Abstract The creation and use of ontologies has become increasingly relevant for complex systems in recent years. This is because of the growing number of use cases that rely on real world integration of disparate systems; the need for semantic congruence across boundaries; and, the expectations of users for conceptual clarity within evolving domains or systems of interest. These needs are evident in most spheres of research involving complex systems but they are especially apparent in infrastructure and cities where traditionally siloed and sectoral approaches have dominated undermining the potential for integration to solve societal challenges such as net zero; resilience to climate change; equity and affordability.

This paper reports on findings of a literature review on infrastructure and cities ontologies and puts forward some hypotheses inferred from the literature findings. The hypotheses are discussed with reference to literature and provide avenues for further research on (1) belief systems that underpin *non* top level ontologies and the potential for interference from them; (2) the need for a small number of top level ontologies and translation mechanisms between them; (3) clarity on the role of standards and information systems upon the adaptability and quality of datasets using ontologies. We also identify a gap in the extent ontologies can support more complex automated coupling and data transformation when dealing with different scales.

1. Introduction

Ontologies in the field of knowledge engineering are sometimes referred to as data models particularly in industry (West, 2011). The term ontology originates in the field of philosophy, where it can be described as the study of what exists, or the study of being (Simons, 2015). Ontology addresses the metaphysical question of “what is there?” Metaphysicians are interested in differentiating the different ways that things can exist, that is, the categories of existence. Some have distinguished concrete objects which exist in space-time from abstract entities that do not. Others have claimed there are no abstract entities (Rosen, 2020) thus it is not surprising to find pervasive pluralism in computational ontologies.

In the research domains of infrastructure and cities a variety of ontologies have been defined. Those with high specificity are mostly linked to specific use cases that address application specific questions often via cyber physical systems using sensors. Abstract entities can exist at all levels (Zhang, Silvescu, & Honavar, 2002) but in infrastructure and cities literature they are usually found in domain, mid or top level ontologies.

In academic literature there are competing ontologies within energy, transport, water, waste and telecoms sectors, as there are for infrastructure and cities. Data sets in practice may be described by meta-data and/or with reference to classification schemes and standards, but this falls short of explicit definition of the structure and nature of the data which could be provided by ontologies. In practice, data sets are regularly implemented without ontological consideration. Without explicit top level ontological commitment it is difficult to: automate reasoning; develop inference (through logic); know the precision of data; differentiate between continuants and occurrents; be certain of data provenance (and connections to the semantic world); and, in general, achieve interoperability (E.g. Leal, Cook, Partridge, Sullivan, & West, 2020)

This paper is organised as follows. The literature review methodology is presented followed by the review findings. The discussion presents three hypotheses deduced from literature and developed further to expose potential gaps and issues. It then considers the practical implications for dealing with different scales. The conclusion presents the avenues for further research and suggestions to extend the scale, scope and methods to reduce the limitations of the work.

2. Methodology

The SCOPUS database was searched for articles in English which include in their title “Ontolog*” and any of the following terms: transport*, road, energy, water, waste, telecom*, 5g, wireless, internet, renewable, smart grid, network, rail, vehicle, shipping, freight, aviation, sewage, treatment, software, cities, infrastructure. The search string is shown in Table 1. Articles outside the scope of economic infrastructure and cities, such as manufacturing, were excluded. 109 articles remained for categorization.

In addition to sectoral findings which are discussed below, the main finding arising from analysis of selected articles was the different levels of ontology. These levels include top (or foundational), mid, domain, application, and sensor. Top level ontologies are especially important for semantic reasoning and integration across domains, yet these are the least well integrated across sectors and domains.

3. Review findings

Given various institutional, regulatory and organizational divisions, it is not surprising that economic (energy, transport, water, waste and telecommunications) infrastructure knowledge is distributed among various disciplines and sectors. For each sector we observed application (purposeful, use case, problem) oriented ontologies as well as domain ontologies (describing entities in a sector). For cyber physical systems we observe sensor ontologies. Some sensor, application, and domain ontologies make a commitment to a top level ontology that is sector agnostic. Top level ontologies provide the generalisations for the structure and organisation of entities: defining different types of entity, how they are related, and allowing for automated reasoning when specific entities appear in lower level ontologies. Figure 1 illustrates the levels of ontology detected.

(Insert Figure 1 about here)

The following sections describe the different types of ontology that appear in literature, exposing their scale and scope. The review is organised naturally by sector, then system-wise for infrastructure and cities. Sectors together constitute infrastructure, so infrastructure ontologies will embrace multiple domain sectors. City ontologies touch all infrastructure sectors and infrastructure as a whole system. Figure 2 illustrates the domain ontology overlaps.

(Insert Figure 2 about here)

3.1 Energy ontologies

With numerous companies involved in the supply and distribution of energy, data arises from a range of sources, for sharing across the sector. A common method in the energy sector is to use device ontologies, particularly SSN ontology (Compton et al., 2012), to bring information together in a common format (Corry, Pauwels, Hu, Keane, & O'Donnell, 2015; Dey, Jaiswal, Dasgupta, &

Mukherjee, 2015). An example for the purpose of smart energy management in buildings is provided by the OPTIMUS ontology (Marinakis & Doukas, 2018). Scale of energy ontologies varies widely from household or building-level energy consumption, to entire cities, districts or urban areas. Perhaps the most extensive example of an urban energy domain ontology is SEMANCO which aims to make urban planning and management more energy efficient. Including urban space descriptors, energy and emission indicators, and socio-economic factors, this is a comprehensive attempt at an energy planning domain ontology, which draws on standards and use cases to ensure it can be applied to a range of scenarios (Madrazo, Sicilia, & Gamboa, 2012). SEMANCO is linked to the SUMO top level ontology, although several other top level ontologies have been used in the energy domain including basic formal ontology (BFO) which is in the final stages of review to become international standard ISO/IEC PRF 21838-2.2[*], Unified Foundation Ontology (UFO) (Guizzardi, Wagner, Almeida, & Guizzardi, 2015), and Business Objects Reference Ontology (BORO) (de Cesare & Partridge, 2016). It is relevant to consider to the objectives of the SEMANCO project which are to foster the use of standards in energy data modelling, to formulate verifiable methods to measure energy performance, to promote the participation of multiple stakeholders in carbon reduction planning, and to provide inputs for future EU policy development[†].

3.2 Water ontologies

Water is perhaps one of the broadest and most difficult domains to define in infrastructure, with the social, economic, and environmental considerations and complexities of the water domain rendering the creation of ontologies in this sector challenging. The vertical integration of potable water distribution and treatment, in contrast to the many companies involved in energy infrastructure, could go some way to explaining the comparative lack of shared knowledge bases, explaining the very few domain ontologies developed for the water domain. Perhaps the broadest ontology attempted in this sector is the water supply ontology 'WatERP' which aims to coordinate the management of supply and demand in order to reduce water usage and associated energy consumption (Varas, 2013). Most water ontologies are application oriented and delimit their

[*] https://www.iso.org/standard/74572.html

[†] http://www.semanco-project.eu/project.htm

scope to be pertinent accordingly, such as disaster risk evaluation to identify the key influences behind urban flooding (Wu, Shen, Wang, & Wu, 2020), identifying and mitigating failures in the water distribution network (Lin, Sedigh, & Hurson, 2012), and water quality management (Ahmedi, Ahmedi, & Jajaga, 2013). Information describing the water bodies themselves, such as rivers, basins and lakes, and the chemical elements that comprise pollutants and other water quality indicators, can be included through the integration of the existing mid-level ontology SWEET (Semantic Web for Earth and Environmental Terminology).

3.3 Transport ontologies

Unlike other infrastructure domains, the transport sector has seen numerous attempts at domain ontologies, albeit varying in scope. This may be because the boundaries for what constitutes a transport network are much clearer than, for instance, the water domain. Such ontologies can span several types of private and public transport systems (Lorenz, Ohlbach, & Yang, 2005), or focus on a particular mode of transportation and associated infrastructure, such as vehicular and road ontologies (Berdier, 2011; Dardailler, 2012). The breadth of work in this field has been explored and analysed in a survey paper by Katsumi and Fox, who surmise that, while no single ontology covers the full high-level taxonomy of the transport domain, the broad scope of the domain is covered, even if not in a high level of granularity, by the collective ontologies surveyed (Katsumi & Fox, 2018). Katsumi and Fox have themselves prepared a transport planning ontology, as part of an ambitious project to develop a suite of ontologies to represent the urban domain (Katsumi & Fox, 2019). In terms of top level ontologies, Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) is commonly cited (Gangemi, Guarino, Masolo, Oltramari, & Schneider, 2002) as is BORO mentioned earlier. Various application oriented ontologies have been developed: to manage and reduce congestion on public roads (Abberley, Gould, Crockett, & Cheng, 2017; Prathilothamai, Marilakshmi, Majeed, & Viswanathan, 2016); road accident identification (Dardailler, 2012); journey planning (Mnasser, Oliveira, Khemaja, & Abed, 2010), and traffic information (Wanichayapong, Pattara-Atikom, & Peachavanish, 2015).

3.4 Telecoms ontologies

The domain of telecoms is somewhat distinctive from other infrastructure sectors in that it includes a significant amount of digital infrastructure, which evolves much more rapidly than much of the physical infrastructure of other sectors. It is perhaps for this reason, that the telecoms domain as a whole has not seen widespread ontology uptake. Some telecoms domain-specific ontological languages were proposed, predating the dominance of OWL2. Network Description Language (NDL) underpins an ontology for describing complex network topologies and technologies(van der Ham, 2010), while an adaptation of OWL has been developed for telecommunication services, Web Ontology Language for services (OWL-S) (Cao, Li, Qiao, & Meng, 2008). Application ontologies in telecoms have focused on specific types of network: to simplify the configuration of 3G wireless networks (Cleary, Danev, & Donoghue, 2005); optical transport networks based on the ITU-T G.805 and G.872 recommendations (Barcelos, Monteiro, Simões, Garcia, & Segatto, 2009); mobile ontologies as part of the SPICE project (Villalonga et al., 2009) and for 'linked data' (Uzun & Küpper, 2012). More ambitious ontologies attempting to address the challenge of semantic interoperability (Qiao, Li, & Chen, 2012) include the Telecommunications Service Domain Ontology (TSDO). As the complexity and heterogeneity of the telecoms networks increases, simplifying approaches have been proposed. The TOUCAN Ontology (ToCo) asserts that all networks are essentially devices with interfaces with which a user can interact, networks of linked devices. By adopting this premise at the core of ToCo, this domain ontology is able to model small-scale networks such as vehicle-to-vehicle networks and smart home devices, as well as large-scale networks such as satellite networks (Q. Zhou, 2018) and hybrid telecommunication networks (Q. Zhou, Gray, & McLaughlin, 2019). Using this notion of networks as systems of devices that may explain the adoption of device ontologies such as the Internet of Things (IoT) ontologies (Steinmetz, Rettberg, Ribeiro, Schroeder, & Pereira, 2018). This shift to a sensor-focused approach has seen device ontologies such as the IoT-Lite applied to digital twins, to support decision making for operational systems (Bermudez-Edo, Elsaleh, Barnaghi, & Taylor, 2015). Taking the concept of device as a starting point, the SAREF ontology for smart appliances (TNO, 2015) has been extended, using GeoSPARQL to represent geospatial data, for the smart city domain (ETSI, 2019). Also well-established is the (OneM2M, 2021) base ontology specifically designed for interoperability for IOT and is built into 4G in the Service Capability Exposure Function (SCEF) function.

3.5 Waste ontologies

Sewage is treated similarly to other linear networks in a sewage ontology as part of an urban description (Heydari, Mansourian, & Taleai, 1991). Perhaps a narrower domain than other infrastructure sectors, the use of ontologies in the waste sector is a relatively new concept. Nonetheless, the field of waste management offers some well-developed ontologies, which have demonstrated their potential through applied case studies, or rule-based reasoning in waste management (Kultsova, Rudnev, Anikin, & Zhukova, 2016). A waste management domain ontology, OntoWM, aligned with the Unified Foundational Ontology (UFO) top level ontology, has been used for monitoring the collection of waste bins and dumpsters (Ahmad, Badr, Salwana, Zakaria, & Tahar, 2018) and can benefit the broader domain of waste management (Sattar, Ahmad, Surin, & Mahmood, 2021). Indeed, as the value of the circular economy model is recognised, the role of waste is shifting from by-product to potential asset. (Trokanas, Cecelja, & Raafat, 2015) created an ontology to represent the domain of industrial symbiosis (IS) (Cecelja et al., 2015). The waste industry is beginning to recognise the importance of knowledge representation in the waste sector. While the use of ontologies remains uncommon, the creation of centralised databases and standards is a valuable step in establishing a solid knowledge base, for example using computer vision and robotics (Recycleye, 2020). Dsposal, the company behind an online platform that links users to a directory of licensed waste facilities, are one of several businesses behind the KnoWaste project, which seeks to connect separate waste systems to achieve greater understanding and enable regulatory oversight. One of the core objectives of the project is the design of an open data standard for waste, on which a central database can be built[‡].

3.6 Infrastructure ontologies

The consequence of sectoral ontologies is that knowledge is not consistently represented across infrastructure. However there have been some attempts to produce an infrastructure domain ontology. The aim is to "*provide an unambiguous formalized representation of domain-wide knowledge in an attempt to provide a shared understanding of domain processes among the various stakeholders for supporting integrated construction and infrastructure development*" [49, p730]. The Infrastructure and Construction PROcess Ontology IC-PRO-Onto aims to serve as a

[‡] https://dsposal.uk/articles/knowaste-govtech-catalyst/ https://github.com/OpenDataManchester/KnoWaste

basis for "*developing further model extensions, domain or application ontologies, software systems, and/or semantic web tools*." (ibid).

3.7 City ontologies

iCity Ontology is an ontology for smart cities (Katsumi & Fox, 2019) building on the Global City Indicators Ontology, which integrates over 10 ontologies from across the semantic web, including geonames, measurement theory, statistics, time, provenance, validity and trust. Elements of existing ontologies have been reused and incorporated where appropriate, including Ontology of Transportation Networks (Lorenz et al., 2005) and Land Based Classification Standards (LBCS) Ontology (Montenegro, Gomes, Urbano, & Duarte, 2012). The iCity project is not aligned to a top level ontology but it leverages the key benefits of working with existing standards. In February 2021, Microsoft launched a Smart Cities Ontology, aligned to their Azure digital twin platform, which utilises ETSI's Application Programming Interface Specification an open framework for context information exchange (ETSI, 2021). Microsoft also make use of ETSI's SAREF extension (Saref4City) in the Smart Cities ontology framework for Topology, Administrative Area and City Object modeling (Russom, Collumbien, De Tant, & Mayrbäurl, 2021).

4. Discussion

Ontologies have the potential for system clarity, exposing biases, overcoming narrow perspectives, rewarding pluralism, and enabling stakeholder engagement. The creation of ontologies itself is a collaborative process with the aim of achieving consensus, identifying gaps, and relying on congruent theories of knowledge. Ontology development can enable discussions on how sustainable, resilient and inclusive outcomes are delivered by integrated engineering systems found in infrastructure and cities.

In order to exploit the potential of ontologies in infrastructure systems and cities, three common threads are identified which are presented as hypotheses and discussed further.

4.1 Toward explicit theories of data-driven ontologies

A preliminary hypothesis H1 is presented in respective of data.

H1: **Ontologies are largely driven from data itself rather than being theory-driven or motivated by normative views**.

Ontologies grounded in data that has been collected in the real world lead to an assumption of value in the data. Thus capturing the data in an ontology creates extra value since ontologies are shareable. Datasets themselves *may be* sufficient to determine their ontology (via some form of inference), however it is much more useful for the developer of a dataset to explicitly define or select the relevant ontology. Indeed, "*Data structures and procedures implicitly or explicitly make commitments to a domain ontology*" [53, p23]. Datasets are assumed to be coherent and rely on theories or belief systems regardless of whether or not they are provided. A dataset commits to a set of things whose existence is acknowledged by a particular theory of system of thought (Partridge et al., 2020). But top level ontologies are different from other ontologies insofar as they do not describe datasets per se. Rather, they define the first order logic of the semantics of the data, or the grammar of the data. Top level ontologies provide the rules that are relevant for semantic interoperability. They need to be formally defined and self-describing. Even the mappings between entities i.e. relationships (which may be: component-to-whole (mereology), set-to-subset (class theory), member-to-class (set theory) and everything else (tuple)) themselves have ontological structure (Purao & Storey, 2005). Furthermore, without explicit top level ontological commitment it is difficult to: automate reasoning; develop inference (through logic); know the precision of data; differentiate between continuants and occurrents; be certain of data provenance (and connections to the semantic world); and, in general, achieve interoperability.

H1 (revised) Non top level ontologies appear not to be theory-driven or motivated by normative views, but are driven by data which fit a belief system which has potential to be inferred.

4.2 Toward a few top level ontologies

A preliminary hypothesis H2 is presented on semantic interoperability.

H2: **Pluralism of top level ontologies within sectors and for infrastructure and cities as whole systems creates a need for translation before semantic interoperability can be delivered.**

Each unique top level ontology represents different ontological commitments such as possibilia, materialism [54, p44] such that formal levels and universal levels of the ontology distinguish top level ontologies from other levels of ontology. Semantic interoperability is not a new concept (Heiler, 1995) and there is a long history of efforts to combine semantic web applications with Building Information Modeling (BIM) and other technologies specific to infrastructure and the built environment (Abanda, Tah, & Keivani, 2013). There is also broad recognition across the built environment of the need to ensure interoperability, which is reflected in standards such as Industry Foundation Classes (IFC) as an interoperable format for BIM data at the building level. Others have attempted ontologies for construction and renovation processes, but many are manually developed despite the need for distributed collaboration among diverse stakeholders and the availability of structured sources (e.g. IFC) as well as unstructured sources (such as safety documents) (Z. Zhou, Goh, & Shen, 2016). In addition to BIM and IFC, (Zhong et al., 2019) identified automated compliance checking through use of ontology and semantic web technology replacing time-consuming, costly and error-prone manual processes especially given then nearly all building projects are modelled digitally.

H2 permits development of options toward semantic interoperability. There is a possibility that some top level ontologies can be abandoned since they have not been defined as thoroughly as others, i.e. they are less complete. However, for different sectors and systems specific top level ontologies appear with wildly varying degrees of adoption. If adoption signifies usefulness it may be possible to eliminate less useful ontologies. For those top level ontologies that remain, it may be possible to create a translation mechanism from one to the other. Using the analogy of language for a top level ontology, experts could translate Latin into Greek. This is simple since these languages do not evolve. For modern languages, translation would require continuous iteration, assuming the translation is even possible. Where translation is found impossible those datasets committing to different top level ontologies cannot be safely integrated. To achieve integration the datasets would need to be reworked to align with an agreed top level ontology.

Thus there two distinct routes forward given H2 and the value of explicit definition of top level ontologies. The first is toward a single infrastructure and cities top level ontology which is usable by all lower level ontologies and their datasets. The second is toward multiple top level ontologies and the development of translation mechanisms between them. Finding a superior top level ontology will be difficult because priorities and perspectives vary and will not be easily reconcilable as a single authority will be needed for decision making.

H2 (revised) **Pluralism of top level ontologies within sectors and for infrastructure and cities as whole systems creates a need to work toward a single top level ontology or a small number of top level ontologies with translation mechanisms before semantic interoperability can be delivered.**

4.3 Toward evolving standards and information management

A preliminary hypothesis H3 is presented on standards.

H3: An adopted standard for a particular data item in more than one infrastructure ontology enables precision for interoperability, however standards can constrain the recognition of emerging values.

Classification systems, taxonomies, standards and other means to organise the potential values of data items provide the means to socialise options and establish validation processes. However this falls short of explicit definition of the structure and nature of the data which could be provided by ontologies. Standards enable assignment of potential types of data (integers, dates, etc.) and enable boundaries to be defined, e.g. not before, not greater than. When standard classification systems such as System International[§] a value of a data item takes on more meaning than without the standard because based on the logic of the standard, the potential of the data can be known. Thus standards have a role to reduce inconsistency between ontologies and associated datasets, however standards have the effect of holding systems in homeostasis due to negative feedback, constraining adaptation. New, legitimate data values can arise in data sets based on the change, reform and nuance required in real-life. When new sub-types of standard values emerge, these are implemented in different ways in ontologies: the standard becomes less precise when new values are admitted, because it is unclear how to process these new values according to the

[§] https://www.npl.co.uk/si-units

standard. Spin-out ontologies or later versions of ontologies can be created which provide for new data items that elaborate the nuances of the adapted item. Regardless changes in data occur more frequently than changes in standards. Standards are reactive.

Another concern on standards is that quality may be incorrectly inferred with respect to other data items in the ontology. Most ontologies will have one or more data items that comply with a standard and very many that do not. The use of one or more important standards may 'rub off' onto the entire ontology undeservedly. Furthermore, the use of an industry standard classification, for example, says nothing about the information management processes used to curate the data. Information management is just as critical as standard selection as it ensures processes of provenance, storage, validation, refresh, etc. are conducted proactively.

> **H3 (revised) An adopted standard for a particular data item in more than one infrastructure ontology enables precision for interoperability, however the use of standards can have unintended and undesirable consequences for adaptation and reasoning about quality.**

4.4 Dealing with scale

Infrastructure research makes use of data, models, conceptualisations and representations of infrastructure systems and linked human, social, economic, political, regulatory, and environmental systems. Objects and processes in each of these systems occur or can be measured, observed or represented over different extents in space and time, and with different levels of detail.

4.4.1 Quantification of scale

The concept of scale relates to orders of magnitude in lengths of space and time, and can be quantified in terms of numerical precision, resolution, extent and coverage. But it also relates to observation and representation of different objects and processes. At the human scale we might be interested in pedestrian flows through stations, where at the catchment scale we look at river flows and reservoir storage.

Reitsma & Bittner (2003) introduce the distinction between extent (spatial size or temporal duration) and granularity (fineness of distinctions or resolution). They consider both endurant objects and perdurant processes to construct an ontological description of scales as ‘hierarchically structured granularity trees’ (ibid:25) where levels in the trees consist of objects or processes of finer granularity and lesser extent as you look further down the tree.

Frank (2009) argues further that domain ontologies are scale-dependent, and observations from remote sensing or sensor networks must include information about their extent and resolution, and that this defines the phenomena that can be represented, giving the example of satellite images which show roads and fields if captured at high resolution, but only patches of field at low resolution.

4.4.2 Scale of representation

The formulation of simulation models and digital twins requires choices to be made about the scale of representation, as well as how to connect models or twins to empirically observed data which may be available with limited extent or resolution again. Multiscale modeling and simulation techniques have been well discussed and developed in computational science and engineering, including in communities of relevance to infrastructure research, in engineering and environmental science (Groen, Zasada, & Coveney, 2014).

Yang & Marquardt (2009) provide an ontological conceptualisation of multiscale modeling. Here scale is used to refer to the multiple levels of abstraction and granularities of representation which are used to model the phenomena of interest, often with reference to numerical principles (finite element decomposition or adaptive meshes) or well-recognised orders of magnitude difference in lengths of space and time (where different physics might be used to model different scales, from quantum mechanics to fluid flow).

Changes in scale of representation are not only a matter of physical sensing and measurement, but also cadastral, administrative and political boundaries and the governance structures that lead to collection of national statistics and surveys. The Office for National Statistics (2019, 2020) posters of the hierarchical representation of UK statistical geographies are an excellent representation of the complexity of simply enumerating the officially-defined sets of areas that are reported against, many of which are updated annually.

4.4.3 Ontological state of scale

Beyond officially-defined geographical extents, there are critical questions of definition and representation of scale. In statistics, the modifiable areal unit problem (Openshaw, 1983) and the ecological fallacy (Gehlke & Biehl, 1934) state the problems of (mis-)representation of spatial phenomena aggregated to different areal units. In human geography, the ontological status of scale has been the subject of debate. Blakey (2020) outlines the moves from theorisations which lean on Kant's understanding of space and time as given, with scales providing a natural ordering and hierarchy, to theories which emphasise politics, power and the social construction of scales (Marston, 2000) and arguments that scales are epistemological and provide contested, various, changing ways of knowing the world that are structured by networks of interaction (Jones, 1998).

The notion of a single natural definition of the extents of cities is also contestable on empirical grounds, as in Arcaute et al., (2015) where a clustering of small areas based on population density and commuting thresholds is used to provide a set of realisations of urban extents in the UK.

4.4.4 Scales in coupled modeling

A software framework for coupling simulation models of infrastructure (smif) is presented in Usher & Russell (2019) along with a brief review of related frameworks, notably the OpenMI standard (Vanecek & Roger, 2014). The smif framework associates the notion of dimensions with model inputs and outputs, where these may be: spatial, comprising a set of areas covering the shared system of interest; temporal, comprising a set of time intervals covering or representing a sample of the shared modelled year; or categorical, where a quantity is represented for multiple categories, such as energy demand by fuel type or economic activity by industrial sector. Following OpenMI conventions, the smif framework introduces adaptors between models when the dimensions of a model output and model input do not match.

Diverse data dimensions produced and required by energy and transport models, such as a subset of the infrastructure simulation models included in NISMOD 2 (ITRC-Mistral, 2020) demonstrate the need to address scale.

4.4.5 Conversion between scales

The methods for converting quantities between dimensions or scales of representation vary according to the phenomenon modelled. For example, energy demand in NISMOD (ITRC-Mistral, 2020) is modelled at Local Authority district regions, with 8760 hourly timesteps (over 365 days) to represent the year. Temperature is an important driver of heating demand and is sampled from a gridded climate model which outputs minimum and maximum temperatures per day.

The energy demand model scales empirically observed demand curves to disaggregate daily minimum and maximum temperatures to get hourly demand for electricity, gas and other fuels for heating. The energy supply model has no notion of demand sectors, so takes demand as the sum across all end uses, and is computationally demanding to run, so samples four representative weeks from the demand time series.

The sampling method aims to preserve the observed peak in demand, which is an important stress test of the power (electricity) supply system, as well as the mean demand for all energy, so that estimates of carbon intensity and total annual generation are consistent with annual demand.

In summary, straightforward aggregation, scaling and proportional disaggregation are sometimes sufficient, sometimes extra information or assumptions are needed to convert values between modelled scales, and sometimes care is needed to preserve particular statistical quantities as values are transformed between scales.

Ontologies for infrastructure research should support the explicit representation and reference to shared definitions of extent and granularity, recognising that definitions change over time, and that datasets and models will use different definitions, so there can be no single preferred scale. Explicit shared definitions are necessary but may not be sufficient to support model coupling and data transformations. Further research could examine to what extent ontologies can support more complex automated coupling and data transformation.

5. Conclusions

This paper reports on findings of a literature review on infrastructure and cities ontologies and puts forward some hypotheses deduced from the literature findings. These hypotheses are discussed with reference to literature and provide avenues for further research on (1) belief systems that underpin *non* top level ontologies and the potential for inference from them; (2) the need for a small number of top level ontologies and translation mechanisms between them; (3) the need for evolving standards and for information systems to improve precision and quality of datasets using ontologies.

These hypotheses underpin practical interventions that are needed to ensure that schemas, metadata, and all scales of representation of data, are organised. The hypotheses must be elaborated and addressed in order for federated digital twins to become a reality. In addition, it is not clear to what extent ontologies can support more complex automated coupling and data transformation when dealing with different scales.

The scope, scale and methods all have limitations which if addressed could materially influence findings. For example, on scope, ontologies could be embraced from construction, buildings, planning, and other aspects of the built environment. On scale, older (and newer) articles could be included, and those with fewer citations or by relaxing quality criteria. On methods, expert opinion especially from the knowledge engineering discipline, use of grey literature, and data modeling expertise for example will add diversity to this academically focussed review.

References (examples only, click here for full details of how to cite your references)

Abanda, F. H., Tah, J. H. M., & Keivani, R. (2013). Trends in built environment Semantic Web applications: Where are we today? *Expert Systems with Applications*, *40*(14), 5563–5577. https://doi.org/10.1016/j.eswa.2013.04.027

Abberley, L., Gould, N., Crockett, K., & Cheng, J. (2017). Modelling road congestion using ontologies for big data analytics in smart cities. *2017 International Smart Cities Conference (ISC2)*, 1–6. https://doi.org/10.1109/ISC2.2017.8090795

Ahmad, M. N., Badr, K. B. A., Salwana, E., Zakaria, N. H., & Tahar, Z. (2018). An Ontology for the Waste Management Domain. *PACIS 2018 Proceedings*, 16. Yokohama, Japan.

Ahmedi, L., Ahmedi, F., & Jajaga, E. (2013, October). *An Ontology Framework for Water Quality Management*. Retrieved from https://www.researchgate.net/publication/255719225_An_Ontology_Framework_for_Water_Quality_Management

Arcaute, E., Hatna, E., Ferguson, P., Youn, H., Johansson, A., & Batty, M. (2015). Constructing cities, deconstructing scaling laws. *Journal of the Royal Society Interface*, *12*(102). https://doi.org/10.1098/rsif.2014.0745

Barcelos, P. P. F., Monteiro, M. E., Simões, R. D. M., Garcia, A. S., & Segatto, M. E. V. (2009). OOTN - An ontology proposal for optical transport networks. *2009 International Conference on Ultra Modern Telecommunications and Workshops*. https://doi.org/10.1109/ICUMT.2009.5345459

Berdier, C. (2011). Road System Ontology: Organisation and Feedback. In G. Falquet, C. Métral, J. Teller, & C. Tweed (Eds.), *Ontologies in Urban Development Projects* (pp. 211–216). https://doi.org/10.1007/978-0-85729-724-2_17

Bermudez-Edo, M., Elsaleh, T., Barnaghi, P., & Taylor, K. (2015). *IoT-Lite Ontology*. Retrieved from https://www.w3.org/Submission/2015/SUBM-iot-lite-20151126/

Blakey, J. (2020). The politics of scale through Rancière. *Progress in Human Geography*. https://doi.org/10.1177/0309132520944487

Cao, D., Li, X., Qiao, X., & Meng, L. (2008). Ontology-based modeling method for semantic telecommunication services. *Proceedings - 5th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2008*, *4*, 449–453. https://doi.org/10.1109/FSKD.2008.283

Cecelja, F., Raafat, T., Trokanas, N., Innes, S., Smith, M., Yang, A., … Kokossis, A. (2015). E-Symbiosis: Technology-enabled support for Industrial Symbiosis targeting Small and Medium Enterprises and innovation. *Journal of Cleaner Production*, *98*, 336–352. https://doi.org/10.1016/j.jclepro.2014.08.051

Chandrasekaran, B., Josephson, J. R., & Benjamins, V. R. (1999). What are Ontologies and why do we need them? *IEEE Intelligent Systems*, *14*(1), 20–26. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=747902

Cleary, D., Danev, B., & Donoghue, D. O. (2005). Using Ontologies to Simplify Wireless Network Configuration. *Proceedings of the 1st International Workshop Formal Ontologies Meet Industry, FOMI 2005*.

Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S., … Taylor, K. (2012). The SSN ontology of the W3C semantic sensor network incubator group. *Journal of Web Semantics*, *17*, 25–32. https://doi.org/10.1016/j.websem.2012.05.003

Corry, E., Pauwels, P., Hu, S., Keane, M., & O'Donnell, J. (2015). A performance assessment ontology for the environmental and energy management of buildings. *Automation in Construction*, *57*, 249–259. https://doi.org/10.1016/j.autcon.2015.05.002

Dardailler, D. (2012). Road Accident Ontology. Retrieved February 7, 2021, from https://www.w3.org/2012/06/rao.html#owl

de Cesare, S., & Partridge, C. (2016). BORO as a Foundation to Enterprise Ontology. *Journal of Information Systems*, *30*(2), 83–112. Retrieved from https://doi.org/10.2308/isys-51428

Dey, S., Jaiswal, D., Dasgupta, R., & Mukherjee, A. (2015). Organization and management of Semantic Sensor information using SSN ontology: An energy meter use case. *2015 9th International Conference on Sensing Technology (ICST)*, 468–473. https://doi.org/10.1109/ICSensT.2015.7438444

El-Gohary, N. M., & El-Diraby, T. E. (2010). Domain Ontology for Processes in Infrastructure and Construction. *Journal of Construction Engineering and Management*, *136*(7), 730–744. https://doi.org/10.1061/(asce)co.1943-7862.0000178

ETSI. (2019). *Extension to SAREF ; Part 4 : Smart Cities Domain* (Vol. 1). Retrieved from https://www.etsi.org/deliver/etsi_ts/103400_103499/10341004/01.01.01_60/ts_10341004v010101p.pdf

ETSI. (2021). Context Information Management (CIM). Retrieved March 9, 2021, from https://www.etsi.org/committee/cim

Frank, A. U. (2009). Geo-Ontologies Are Scale Dependent. *Assembly*, *11*(2005), 13623. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.614.4310&rep=rep1&type=pdf

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., & Schneider, L. (2002). Sweetening Ontologies with DOLCE. In A. Gómez-Pérez & V. R. Benjamins (Eds.), *Knowledge*

Engineering and Knowledge Management: Ontologies and the Semantic Web (pp. 166–181). https://doi.org/10.1007/3-540-45810-7_18

Gehlke, C., & Biehl, K. (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association*, *29*, 169–170.

Groen, D., Zasada, S. J., & Coveney, P. V. (2014). Survey of multiscale and multiphysics applications and communities. *Computing in Science and Engineering*, *16*(2), 34–43. https://doi.org/10.1109/MCSE.2013.47

Guizzardi, G., Wagner, G., Almeida, J. P. A., & Guizzardi, R. S. S. (2015). Towards Ontological Foundations for Conceptual Modeling: The Unified Foundational Ontology (UFO) Story. *Applied Ontology*, *10*(3–4), 259–271.

Heiler, S. (1995). Semantic Interoperability. *ACM Computing Surveys*, *27*(2), 271–273. https://doi.org/10.2307/417557

Heydari, N., Mansourian, a, & Taleai, M. (1991). Ontology-based GIS web service for increasing semantic interoperability among organizations involving drilling in city of Tehran. *Gsdi.Org*, (June 2014). Retrieved from http://www.gsdi.org/gsdi11/papers/pdf/163.pdf

ISO/IEC PRF 21838-2.2. (2021). Retrieved February 22, 2021, from https://www.iso.org/standard/74572.html

ITRC-Mistral. (2020). NISMOD: National Infrastructure Systems MODel. Retrieved March 3, 2021, from https://www.itrc.org.uk/nismod/

Jones, K. T. (1998). Scale as epistemology. *Political Geography*, *17*(1), 25–28. https://doi.org/10.1016/S0962-6298(97)00049-8

Katsumi, M., & Fox, M. (2018). Ontologies for Transportation Research: A Survey. *Transportation Research Part C: Emerging Technologies*, *89*, 53–82. https://doi.org/10.1016/j.trc.2018.01.023

Katsumi, M., & Fox, M. (2019). An ontology-based standard for transportation planning. *CEUR Workshop Proceedings*, *2518*.

Kultsova, M., Rudnev, R., Anikin, A., & Zhukova, I. (2016). An ontology-based approach to intelligent support of decision making in waste management. *IISA 2016 - 7th International*

Conference on Information, Intelligence, Systems and Applications, 2–7. https://doi.org/10.1109/IISA.2016.7785401

Leal, D., Cook, A., Partridge, C., Sullivan, J., & West, M. (2020). *A Survey of Industry Data Models and Reference Data Libraries - to identify requirements for, and provide input to, a Foundation Data Model*. https://doi.org/10.1061/9780784480823.017.

Lin, J., Sedigh, S., & Hurson, A. R. (2012). Ontologies and Decision Support for Failure Mitigation in Intelligent Water Distribution Networks. *2012 45th Annual Hawaii International Conference on System Sciences (HICSS)*, 1187–1196. https://doi.org/10.1109/HICSS.2012.458

Lorenz, B., Ohlbach, H. J., & Yang, L. (2005). *Ontology of Transportation Networks*. Reasoning on the Web with Rules and Semantics.

Madrazo, L., Sicilia, A., & Gamboa, G. (2012). SEMANCO: Semantic Tools for Carbon Reduction in Urban Planning. *3rd Workshop on EeBuildings Data Models (Energy Efficiency Vocabularies)*, 19. Reykjavik.

Marinakis, V., & Doukas, H. (2018). An Advanced IoT-based System for Intelligent Energy Management in Buildings. *Sensors*, *18*(2), 610. https://doi.org/10.3390/s18020610

Marston, S. A. (2000). The social construction of scale. *Progress in Human Geography*, *24*(2), 219–242. https://doi.org/10.1191/030913200674086272

Mnasser, H., Oliveira, K., Khemaja, M., & Abed, M. (2010). Towards an Ontology-based Transportation System for User Travel Planning. *IFAC Proceedings Volumes*, *43*(8), 604–611. https://doi.org/10.3182/20100712-3-FR-2020.00098

Montenegro, N., Gomes, J. C., Urbano, P., & Duarte, J. P. (2012). A Land Use Planning Ontology: LBCS. *Future Internet*, *4*(1), 65–82. https://doi.org/10.3390/fi4010065

OneM2M. (2021). Ontology for M2M. Retrieved February 22, 2021, from https://www.onem2m.org/technical/onem2m-ontologies

ONS. (2019). (Office for National Statistics) Hierarchical Representation of UK Statistical Geographies. Retrieved February 26, 2021, from https://data.gov.uk/dataset/4323fa3d-e2fd-4927-845e-cd80231d21df/hierarchical-representation-of-uk-statistical-geographies-december-2019

ONS. (2020). Office for National Statistics (2020) Hierarchical Representation of UK Statistical

Geographies. Retrieved February 26, 2021, from https://data.gov.uk/dataset/27d3ae04-8e75-43e0-808b-35123345572f/hierarchical-representation-of-uk-statistical-geographies-april-2020

Openshaw, S. (1983). *The modifiable areal unit problem*. Norwich: GeoBooks.

Partridge, C., Mitchell, A., Cook, A., Leal, D., Sullivan, J., & West, M. (2020). *A survey of Top-Level Ontologies*. Retrieved from https://www.cdbb.cam.ac.uk/files/a_survey_of_top-level_ontologies_lowres.pdf

Prathilothamai, M., Marilakshmi, S., Majeed, N., & Viswanathan, V. (2016). Timely Prediction of Road Traffic Congestion Using Ontology. In L. P. Suresh & B. K. Panigrahi (Eds.), *Proceedings of the International Conference on Soft Computing Systems* (pp. 331–344). https://doi.org/10.1007/978-81-322-2674-1_32

Purao, S., & Storey, V. C. (2005). A multi-layered ontology for comparing relationship semantics in conceptual models of databases. *Applied Ontology*, *1*(March), 117–139.

Qiao, X., Li, X., & Chen, J. (2012). Telecommunications Service Domain Ontology: Semantic Interoperation Foundation of Intelligent Integrated Services. In *Telecommunications Networks - Current Status and Future Trends* (pp. 183–210). https://doi.org/DOI: 10.5772/36794

Recycleye. (2020). WasteNet: The World's Largest Dataset for Waste. Retrieved March 2, 2021, from https://recycleye.com/wastenet/

Reitsma, F., & Bittner, T. (2003). Scale in Object and Process Ontologies. In W. Kuhn, M. Warboys, & S. Timpf (Eds.), *Spatial Information Theory. Foundations of Geographic Information Science, Lecture Notes in Computer Science* (pp. 13–27). Retrieved from https://doi.org/10.1007/978-3-540-39923-0_2

Rosen, G. (2020). Abstract Objects. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Spring 2020)*. Retrieved from https://plato.stanford.edu/archives/spr2020/entries/abstract-objects/

Russom, M. B., Collumbien, A., De Tant, G., & Mayrbäurl, J. (2021). Azure/opendigitaltwins-smartcities. Retrieved March 9, 2021, from https://github.com/Azure/opendigitaltwins-smartcities/

Sattar, A., Ahmad, M. N., Surin, E. S. M., & Mahmood, A. K. (2021). An improved methodology for collaborative construction of reusable, localized, and shareable ontology. *IEEE Access*, *9*, 17463–17484. https://doi.org/10.1109/ACCESS.2021.3054412

Simons, P. M. (2015, January). Ontology. *Encyclopedia Britannica*. Retrieved from https://www.britannica.com/topic/ontology-metaphysics

Steinmetz, C., Rettberg, A., Ribeiro, F. G. C., Schroeder, G., & Pereira, C. E. (2018). Internet of things ontology for digital twin in cyber physical systems. *Brazilian Symposium on Computing System Engineering, SBESC*, *2018-Novem*, 154–159. https://doi.org/10.1109/SBESC.2018.00030

TNO. (2015). Smart Appliances Project. Retrieved March 2, 2021, from https://sites.google.com/site/smartappliancesproject/home

Trokanas, N., Cecelja, F., & Raafat, T. (2015). Semantic approach for pre-assessment of environmental indicators in Industrial Symbiosis. *Journal of Cleaner Production*, *96*, 349–361. https://doi.org/10.1016/j.jclepro.2013.12.046

Usher, W., & Russell, T. (2019). A software framework for the integration of infrastructure simulation models. *Journal of Open Research Software*, *7*(1), 16. https://doi.org/10.5334/jors.265

Uzun, A., & Küpper, A. (2012). OpenMobileNetwork - Extending the Web of Data by a dataset for mobile networks and devices. *ACM International Conference Proceeding Series*, 17–24. https://doi.org/10.1145/2362499.2362503

van der Ham, J. (2010). *A semantic model for complex computer networks: the network description language*. Retrieved from https://jvdham.nl/publication/vanderham-2010-318784/vanderHam-2010-318784.pdf

Vanecek, S., & Roger, M. (2014). OGC® Open Modelling Interface Interface Standard. Retrieved February 26, 2021, from OGC 07-041r1 website: http://www.opengeospatial.org/standards/openmi

Varas, G. A. (2013). *Generic ontology for water supply distribution chain*. Retrieved from https://cordis.europa.eu/docs/projects/cnect/3/318603/080/deliverables/001-D13GenericOntologyforwatersupplydistributionchainv13.pdf

Villalonga, C., Strohbach, M., Snoeck, N., Sutterer, M., Belaunde, M., Kovacs, E., … Droegehorn, O. (2009). Mobile Ontology: Towards a Standardized Semantic Model for the

Mobile Domain. *Service-Oriented Computing - ICSOC 2007 Workshops*, 248–257. Retrieved from https://link.springer.com/chapter/10.1007/978-3-540-93851-4_25

Wanichayapong, N., Pattara-Atikom, W., & Peachavanish, R. (2015). Road Traffic Question Answering System Using Ontology. In T. Supnithi, T. Yamaguchi, J. Z. Pan, V. Wuwongse, & M. Buranarach (Eds.), *Semantic Technology* (pp. 422–427). https://doi.org/10.1007/978-3-319-15615-6_32

West, M. (2011). *Developing High Quality Data Models*. Retrieved from https://www.sciencedirect.com/science/article/pii/B9780123751065000221

Wu, Z., Shen, Y., Wang, H., & Wu, M. (2020). Urban flood disaster risk evaluation based on ontology and Bayesian Network. *Journal of Hydrology*, *583*, 124596. https://doi.org/10.1016/j.jhydrol.2020.124596

Yang, A., & Marquardt, W. (2009). An ontological conceptualization of multiscale models. *Computers and Chemical Engineering*, *33*(4), 822–837. https://doi.org/10.1016/j.compchemeng.2008.11.015

Zhang, J., Silvescu, A., & Honavar, V. (2002). Ontology-Driven Induction of Decision Trees at Multiple Levels of Abstraction. In *Abstraction, Reformulation and Approximation* (pp. 316, 323). Retrieved from https://link.springer.com/content/pdf/10.1007%2F3-540-45622-8.pdf

Zhong, B., Wu, H., Li, H., Sepasgozar, S., Luo, H., & He, L. (2019). A scientometric analysis and critical review of construction related ontology research. *Automation in Construction*, *101*(June 2018), 17–31. https://doi.org/10.1016/j.autcon.2018.12.013

Zhou, Q. (2018). *Ontology-Driven Knowledge Based Autonomic Management for Telecommunication Networks : Theory , Implementation , and Applications* (Heriot-Watt University, Edinburgh, UK). Retrieved from https://www.ros.hw.ac.uk/handle/10399/3497

Zhou, Q., Gray, A. J. G., & McLaughlin, S. (2019). ToCo: An ontology for representing hybrid telecommunication networks. In *The Semantic Web: ESWC 2019. Lecture Notes in Computer Science, 16th Extended Semantic Web Conference, Portorož, Slovenia*: *Vol. 11503 LNCS* (pp. 507–522). https://doi.org/10.1007/978-3-030-21348-0_33

Zhou, Z., Goh, Y. M., & Shen, L. (2016). Overview and Analysis of Ontology Studies Supporting Development of the Construction Industry. *Journal of Computing in Civil Engineering*, *30*(6), 04016026. https://doi.org/10.1061/(asce)cp.1943-5487.0000594