

INFRASTRUCTURE RESEARCH ONTOLOGIES

FINAL REPORT

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

31.03.2021 Final

Table of Figures	3
Table of Tables	3
Executive Summary	4
1 Introduction	5
1.1 Rationale	5
1.1 Project aims and report structure	6
1.2 Project contributors and parallel developments.....	7
2 Methodology	8
3 State of the Art.....	9
3.1 Definitions of Ontologies	9
3.2 Principles for ecosystems using ontologies	10
3.3 Levels of ontology.....	10
3.4 Domain (sectoral) ontologies	14
3.5 Infrastructure research dealing with scale	20
3.6 Ontologies of ontologies	23
3.7 Ontology languages and Resource Descriptors	25
3.8 Literature on tooling, technologies, and methods.....	25
4 Industrial Practices.....	30
4.1 Data and Analytics Facility for National Infrastructure (DAFNI) data	30
4.2 Industrial Ontologies	32
4.3 Mid level ontologies	37
4.4 Barriers, risks, opportunities and strategies	40
4.5 Digital Twins: capabilities and governance.....	42
5 Practitioners Panel Discussion	45
5.1 Interoperability	45
5.2 Knowledge and evolution	51
5.3 Use cases.....	53
6 Recommendations, assumptions and next steps	58
6.1 Recommendations	58
6.2 Assumptions	58
6.3 Next steps.....	61
6.4 Gaps and limitations	63
7 Acknowledgements.....	64
8 References	65

Table of Figures

Figure 1. Ontology levels and their generalisability	12
Figure 2. Venn diagram of infrastructure ontologies	12
Figure 3. Venn diagram of infrastructure ontologies overlaid with ontologies from energy sector	15
Figure 4. Venn diagram of infrastructure ontologies overlaid with ontologies from water sector	16
Figure 5. Venn diagram of infrastructure ontologies overlaid with ontologies from transport sector	18
Figure 6. Diverse data dimensions across infrastructure. Usher & Russell (2019).....	22
Figure 7. The architecture of the IOF ontologies (Industrial Ontologies Foundry, 2021)	33
Figure 8. Industry data models and reference data libraries by application area (Leal et al., 2020)	34
Figure 9. Equinor ontology based mediation between data users and data sources (Kharlamov et al., 2019)	35
Figure 10. The SOMM editor.....	36
Figure 11. The terminology used by Highways England business units	36
Figure 12. the iCity ontology (Katsumi & Fox, 2019)	37
Figure 13. Digital Twin ontology (Singh et al., 2020)	38
Figure 14. Ontology functional mapping (Singh et al., 2020)	38
Figure 15. Azure Smart Cities Ontology for Digital Twins (Russom et al., 2021)	39
Figure 16. Capabilities to make progress on an ecosystem of digital twins	43
Figure 17. Approaches toward trialling possible solutions	43
Figure 18: Governance models for digital twin ecosystem.....	44
Figure 19. Knowledge representation paradigms (LIU & ÖZSU, 2009)	51
Figure 20. Ontology development (Žáček, 2017).....	52
Figure 21. Roadmap showing mandatory and optional standards for autonomous vehicle and intelligent transport system © John Batterbee, Costain	55

Table of Tables

Table 1. SCOPUS search string to identify literature	11
Table 2: Sectoral barriers, opportunities, risks and strategies.....	41
Table 3: Use case options.....	62

Executive Summary

The purpose of the IRO project was to

1. make recommendations to progress toward a top level infrastructure research ontology
2. identify barriers, gaps and priority areas
3. provide a shared framing for stakeholder engagement.

In order to meet its aims, the research design melds both academic and industrial knowledge to directly engage with the ontologies research community to provide recommendations for developing this field of study.

State of the art academic literature is reviewed first in chapter 3, discovering an array of ontologies from different infrastructure sectors. Ontologies may be usefully organized by level, the top level ontologies being at the foundation level and being very distinct in nature from other ontologies. Current methods and approaches for dealing with different scales (e.g. temporal, spatial scales) in infrastructure research highlight cross-sectoral differences which ontologies can assist to reconcile. Methods to integrate ontologies, tools to describe and manipulate ontologies and methods for manipulating datasets demonstrate some clearly emerging good practices.

Industrial practices are also reviewed and presented in chapter 4. DAFNI, the Data Analytics Facility for National Infrastructure, which is to be the primary location for data, models and visualisations for infrastructure research is examined first in terms of its approach to data management. Key ontologies from infrastructure industries are then identified, followed by mid level ontologies which are attempts to conjoin top level and domain ontologies, often across broad domains such as cities, and digital twins. Data collected from stakeholders and practitioners focused on barriers, risks, and opportunities for ontologies in the context of digital twins in infrastructure systems. Stakeholders were also asked to identify preferred modes of governance, trialling and developing capabilities.

These insights were brought together and presented to the ontology and digital twin communities working in infrastructure and the built environment in chapter 5 **on Practitioners Panel Discussion**. Opinions were sought and clarified particularly on the role of a single top level ontologies, as this has the greatest integrating potential. Discussions led to reflection on the entire knowledge engineering industry, and enabled the project use case, on autonomous vehicles, to be addressed.

The findings demonstrate a substantial variety of ontologies and approaches in infrastructure sectors and cyber physical systems (or Internet of Things). There is no single top level ontology that currently enables their integration. Furthermore, the fast paced development of digital twins is likely to mean a plethora of new models based on existing technological approaches, rather than using a constructional ontology-led approach enabling whole systems plurality. The findings were reviewed and brought together in chapter 6, as a set of recommendations. In discovering these recommendations, it was useful to tease out the underlying assumptions of the project team, to expose biases and preferences. A set of actions for next steps were developed:

- the creation of demonstrators as proofs of concept and to engage with stakeholders,
- action on DAFNI to prepare for ontologies and digital twins
- the development of the network of collaborators, and
- the creation of new governance and business models.

1 Introduction

This is the final report of the DAFNI funded champion project called 'Infrastructure Research Ontologies' (IRO).

1.1 Rationale

Infrastructure research is the process of creating insight about infrastructure systems. For our purposes, infrastructure is the system of systems of all transport, energy, water, waste and telecommunications systems, sometimes called economic infrastructure (HM Infrastructure and Projects Authority, 2016). Infrastructure provides services such as mobility, heating, power, potable water, and information which are critical not just to society but to the economy. The choices made for the design, construction, operation and decommissioning of infrastructure create long-term environmental impacts affecting the sustainability of infrastructure, and also public health. Given also concerns on climate change, and other threats, natural and man-made, which are leading to more frequent and severe conditions and uncertainties, new methods and capabilities are needed to create robustness, reliability and resilience of infrastructure. Infrastructure research embraces these integrated challenges and increasingly closer to real-time data is creating an excellent opportunity to avoid disasters, reduce environmental impact, increase efficiency, and improve public health, leading to better social and economic outcomes.

The processes used to create insight invariably involve quantitative methods and digitalized information. There are hundreds, if not thousands, of models and data sets used in infrastructure research. Yet there is inconsistency when attempting to match data when a data set in one infrastructure sector using a different mechanism to conceptualise a 'thing' than another sector. A particularly good example of this is scale, where different sectors measure 'things' over different temporal and spatial scales. Different sectors, even sub-sectors and specific organisations, also devise their own taxonomies, or classifications, to describe their data entities and relationships. In the past, this may have worked to a large extent, but with the development of Industry 4.0, and the related technological capability it is set to provide, infrastructure needs to be considered holistically, certainly in terms of the data each part of it will need to share. Digital twins, sometimes known as synthetic environments, are processes and technologies that collect real world data from diverse sources enabled by digital representations, intelligently bringing them together with digital technology and providing insight for proactive interventions, in a circular flow of information.

Addressing the integration of data that resolves existing idiosyncrasies elevates the importance of information management and the need to find innovative ways of integrating data across infrastructure systems. The scope of information management encompasses both the mechanisms for organizing and sharing information, as well as the representations of information at abstract (or conceptual) level and data level, as well core approaches to understanding information (ways of knowing and describing). The work here pays particular attention to the diversity of conceptual descriptions (ontologies) across infrastructure systems, in both academia and industry. It addresses a gap in understanding on the nature and role of ontologies for infrastructure research which must be addressed for future research to be successful and to avoid it taking narrow and partial insights. The work highlights the presence of multiple ways of knowing (top level ontologies) even within single sectors of infrastructure research. This plurality is perceived to be a barrier to accurate, appropriate and trusted integration between disparate data and models of infrastructure systems. Moreover, the absence of an explicit top level ontology is a concern in the context of the emerging trend for digital twins, since without the top level there is no integration by design: integration cannot

be robust without the structural support of a top level ontology providing an intelligent integrative framework.

The Grenfell disaster is a case in point in the built environment whereby diverse digital records of buildings with particular characteristics were not readily available in an integrated way facilitating their collection, the selection of relevant information within them and the timely provision of consistent assessment. There is a clear need to be proactive in infrastructure systems and the built environment more generally to ensure public health, safety and security, and prioritizing interventions to address risks and uncertainties via access to relevant information at different scales.

1.1 Project aims and report structure

The purpose of the IRO project is to

4. make recommendations to progress toward a top level infrastructure research ontology
5. identify barriers, gaps and priority areas
6. provide a shared framing for stakeholder engagement.

In order to meet its aims, the research design melds both academic and industrial knowledge to directly engage with the ontologies research community to provide recommendations for developing this field of study.

State of the art academic literature is reviewed first in chapter 3, discovering an array of ontologies from different infrastructure sectors. Ontologies may be usefully organized by level, the top level ontologies being at the foundation level and being very distinct in nature from other ontologies. Current methods and approaches for dealing with different scales (e.g. temporal, spatial scales) in infrastructure research highlight cross-sectoral differences which ontologies can assist to reconcile. Methods to integrate ontologies, tools to describe and manipulate ontologies and methods for manipulating datasets demonstrate some clearly emerging good practices.

Industrial practices are also reviewed and presented in chapter 4. DAFNI, the Data Analytics Facility for National Infrastructure, which is to be the primary location for data, models and visualisations for infrastructure research is examined first in terms of its approach to data management. Key ontologies from infrastructure industries are then identified, followed by mid level ontologies which are attempts to conjoin top level and domain ontologies, often across broad domains such as cities, and digital twins. Data collected from stakeholders and practitioners focused on barriers, risks, and opportunities for ontologies in the context of digital twins in infrastructure systems. Stakeholders were also asked to identify preferred modes of governance, trialling and developing capabilities.

These insights were brought together and presented to the ontology and digital twin communities working in infrastructure and the built environment in chapter 5 **on Practitioners Panel Discussion**. Opinions were sought and clarified particularly on the role of a single top level ontologies, as this has the greatest integrating potential. Discussions led to reflection on the entire knowledge engineering industry, and enabled the project use case, on autonomous vehicles, to be addressed.

The findings were reviewed and brought together in chapter 6, as a set of recommendations. In discovering these recommendations, it was useful to tease out the underlying assumptions of the project team, to expose biases and preferences. A set of actions for next steps are developed which include action on DAFNI to prepare for ontologies and digital twins, the creation of demonstrators as proof of concept and to engage with stakeholders, the development of the network of collaborators, and the creation of new governance and business models. Gaps and limitations are then discussed.

1.2 Project contributors and parallel developments

The work was a collaboration between four universities led by UCL, and with Cranfield, Oxford and Newcastle Universities, the STFC, and with two large companies: Costain and Arup. Each university had specific responsibilities for themes of investigations: infrastructure ontologies literature, industrial ontologies and stakeholder engagement (UCL); DAFNI data and taxonomies (Cranfield and STFC); scale (Oxford); ontology of ontologies, methods and tools (Newcastle).

In parallel the Centre for Digital Built Britain at Cambridge University has been working on a National Digital Twin (NDT) programme. They have set out the technical approach for the development of an Information Management Framework (IMF) for the built environment (Hetherington & West, 2020). The purpose of the IMF is to enable secure, resilient data sharing across the built environment. CDBB have also been examining top level and industrial ontologies for the built environment (Partridge et al., 2020).

Authors of the CDBB reports are consultants and members of the 4DSIG (a special interest group on 4D ontologies) led by Dame Professor Wendy Hall at Southampton University. 4DSIG members have provided an excellent source of both innovation and validation and we acknowledge their insights and feedback. Furthermore members of the team working on the Digital Twin Gemini project “Digital Twin Toolkit” (Hayes, 2021) (to which two of the co-authors of this report contributed) were also an excellent source of challenge and inspiration.

2 Methodology

Desk-based methods using on-line databases were a primary method for the state of the art review and to identify industrial practices.

A stakeholder network was developed by identifying and contacting contributors to articles, authors of methods, leaders of practices, and through snow-balling of these, and project team contacts. Once key players were identified, they were approached for interview.

Working under Covid-19 lock-down, all progress was electronic, however we managed to hold two large on-line dissemination events on ontologies for digital twins. The events included stakeholder participation (via Chat, oral discussion and use of Mural board – on-line white board).

The first event included representatives and speakers across industry transport, energy and water sectors. We obtained responses to poll questions from over 60 respondents and collected 16 detailed responses to an ontologies survey.

The second included specialists in ontology, including academics, and consultants, who work in areas of security, and national interest. These provided material for discussions for Section 5 and were further verified with Dr Matthew West, OBE.

The work was enriched by having both Costain and Arup in the project team: both are leading on digital innovation, and Costain kindly set the use case challenge for the project. Otherwise many thanks go to those we acknowledge (see after conclusions) who provided freely of their time and excellent insights.

The Cranfield University work on data sets and taxonomies is available on the DAFNI web site and contributes to the overall discussion and conclusions.

Slack and Teams were used for project management and communication, providing the means for the sharing of discoveries and debates to be conducted.

3 State of the Art

3.1 Definitions of Ontologies

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). More radically, some have sanctioned a kind of existence (subsistence) of impossible objects (Yagisawa, 2020). Said another way, immaterial things “subsist” rather than “exist”: respectively, there are negative existentials (things that don’t exist) and universals (things in common: characteristics or qualities) (Quine, 1980).

A standard description of ontology is “*the set of things whose existence is acknowledged by a particular theory or system of thought*” (Lowe, 1995). A particular instance of an ontology will select a sub-set of information which takes a particular perspective or has theoretical coherence. This arises since not all information about either an existent or non-existent object is needed (or available, or usable, etc.) for the purpose of defining a specific ontology. Indeed choices about what information to include for the specific task at hand are often made without robust ontological practices, i.e. the information is just gathered into data sets without explicit concern for ontological coherence.

The objects that a system or dataset acknowledgement is called its ontic commitment. This can be regarded as the ontological cost (Bricker, 2016). So the ontological cost, regardless of whether the ontology is defined explicitly or not, can be calculated and plays a polemical role: one theory could be argued as more costly than another.

Thus, philosophical understanding of ontologies provides a framing for ontologies of systems and datasets This is the primary concern of this paper, since it concerns top-level or foundational ontologies, and theoretically provides a way forward for semantic interoperability. Assuming that the semantics for specific datasets can be resolved by their ontologies, then a top-level ontology can resolve disparate systems of thought, thus it can integrate them.

Where there is an interest in reuse and interoperability, the top categorical level becomes particularly relevant as it helps to ensure semantic consistency and coherence. Often the general categories of existence are not explicitly acknowledged in the system, and these need to be re-engineered. Understandably, this re-engineering will raise questions of consistency and coherence.

Translation can occur between ontologies and avoids costly peer to peer integration between datasets. Mapping to a single common foundational or Top Level Ontology (TLO) is the cheapest option. Furthermore, new datasets can adopt the TLO avoiding any mapping needs.

The integration capability of a top level ontology provides the cheapest mechanism for sharing knowledge across diverse domains.

3.2 Principles for ecosystems using ontologies

The sub-set of information that is contained in an ontology (whether explicitly stated or not in respect of a dataset) is determined by the principles of the ecosystems in which the datasets are used. For example, if there is a principle regarding security of information, then the dataset will contain objects that enable security to be implemented. The implementation of principles is constrained by the operational limitations of the systems in which the datasets are used, and the digital maturity of technological solutions. For example, security may be implemented by identity checking, which can be solved by simple password mechanisms or more robustly by biometric means.

The Gemini Principles (Bolton, Enzer, & Schooling, 2018) for digital twins set out nine principles to align approaches to information management across the built environment: Purpose (Public good, Value creation, Insight); Trust (Security, Openness, Quality); and Function (Federation, Curation, Evolution). The principles are system qualities or non-functional requirements acting as a ‘conscience’ (Gerber, Nguyen, & Gaetani, 2019, p62) for doing the right or ethical things.

An information management framework (the *how*) together with a top level ontology (and multitude lower level ontologies) (the *what*) determine the nature (or characteristics) of federated digital solutions. For example, “*core security principles will need to be carefully considered and a proper security architecture put in place to develop safe and secure digital twins*” (Gerber et al., 2019, p41). In strong support of a security principle others (4DSIG, 2020) advocate the holistic representation of security characteristics of objects, including: software vulnerabilities (where known), security enforcing functionality, assurance of cyber-physical, cyber and backend systems, and mechanisms for resistance.

Another example, is the principle of trust (regarding ethical aspects of both data and technologies) which must be addressed before widespread deployment of cyber-physical (Internet of Things) deployment (Gerber et al., 2019, P50). Security and trust (as do other principles) demand attention in both the Information Management Framework and in ontologies in order to have successful federation of digital twins, including where does/ are/ is the rights, portability/ mobility, consent, etc. (4DSIG, 2020). Without a consistent approach to what is represented in data it is very hard to formalise principles.

Explicit identification of the core principles of ecosystems in which ontologies will be used provides direction, improves clarity, and facilitates consistency.
--

3.3 Levels of ontology

3.3.1 Literature identification

Academic literature was discovered using the SCOPUS database by applying search strings as shown in Table 1. Articles on the intersection of ontologies for infrastructure was the target however the search results required filtering to exclude articles outside scope, e.g. to exclude manufacturing articles. 109 articles remained for categorization. The levels of ontology was the clearest category and are discussed in this section, although a sectoral categorization is provided in section 3.4.

Table 1. SCOPUS search string to identify literature

```
TITLE ("Ontolog*" AND ( transport* OR road OR energy OR water OR waste OR
telecom* OR telecommunication* OR 5g OR wireless OR internet OR renewable
OR ( smart AND grid ) OR network OR rail OR vehicle OR shipping OR freight
OR aviation OR sewage OR treatment OR software )) AND ( LIMIT-TO (
LANGUAGE , "English" )) AND ( EXCLUDE ( SUBJAREA , "COMP" ) OR EXCLUDE (
SUBJAREA , "MATH" ) OR EXCLUDE ( SUBJAREA , "SOCI" ) OR EXCLUDE (
SUBJAREA , "DECI" ) OR EXCLUDE ( SUBJAREA , "BIOC" ) OR EXCLUDE (
SUBJAREA , "BUSI" ) OR EXCLUDE ( SUBJAREA , "MEDI" ) OR EXCLUDE (
SUBJAREA , "ARTS" ) OR EXCLUDE ( SUBJAREA , "PHYS" ) OR EXCLUDE (
SUBJAREA , "MATE" ) OR EXCLUDE ( SUBJAREA , "CENG" ) OR EXCLUDE (
SUBJAREA , "AGRI" ) OR EXCLUDE ( SUBJAREA , "HEAL" ) OR EXCLUDE (
SUBJAREA , "CHEM" ) OR EXCLUDE ( SUBJAREA , "PHAR" ) OR EXCLUDE (
SUBJAREA , "IMMU" ) OR EXCLUDE ( SUBJAREA , "PSYC" ) OR EXCLUDE (
SUBJAREA , "NEUR" ) OR EXCLUDE ( SUBJAREA , "ECON" ) OR EXCLUDE (
SUBJAREA , "NURS" ) )
```

3.3.2 Overview

Individual ontologies for knowledge engineering in the discipline of computer science can sit anywhere within a broad spectrum of abstraction, and vary significantly based on their intended purpose. Five main types of ontology have been identified as relevant for infrastructure research. These are presented in Figure 1.

- **Top level ontologies** define the general structure of concepts and relationships (what there is and how it is related). This is a significant topic for this report and is discussed in more detail in section 5.
- **Mid level ontologies** bridge between top level and domain level ontologies. They do not however provide the explicit statements on ontological commitment which can be found in top level ontologies.
- **Device ontologies** describe sensor, actuator and device concepts relevant for cyber-physical systems (Internet of Things). Sensors and controllers have a large part to play in digital twins, since they are means of obtaining data and actuating interventions into real world systems. Devices are often domain independent.
- **Domain ontologies** describe the system of interest, which may be a sector, or a nexus, such as a water-energy-food system.
- **Applied ontologies** are application oriented: for problem-solving or use-case.

Top level ontologies are the most generalisable (and useful for defining information relevant to lower level ontologies) whereas application ontologies are the most specific (and draw on higher level ontologies). W3C's Semantic Sensor Network (SSN) Ontology (Atkinson, García-Castro, Lieberman, & Stadler, 2017) SSN has been incorporated into domain (including cross-sectoral domain) so it is positioned at a level 'higher' than domain in the level of ontologies. An alternative lens may perceive it as a specialised domain or even a vertical component of a mid level ontology for the modern infrastructure domains.

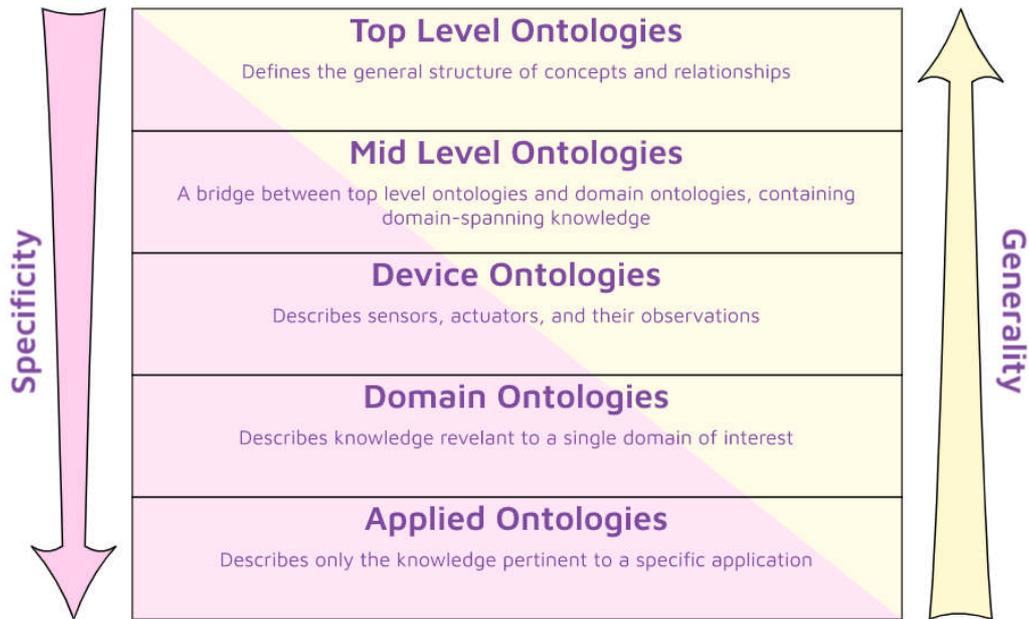


Figure 1. Ontology levels and their generalisability

Whilst these levels are useful for classification purposes, there are examples in literature where different ontology levels are accommodated into one ontology. For example, the Geographic Data Files (GDF) standard has been formalised into a GDF ontology for **transport** network use. It declares three levels within one ontology (Lorenz, 2005):

- Level 0 (Topology): which describes the fundamental geometrical and topological entities;
- Level 1 (Features): which adds the possibility to describe real world geographic objects with their characterising properties;
- Level 2: which gives the possibility to describe complex features which are aggregates of other features

In the waste sector, Trokanas, Cecelja, & Raafat (2014) develop an industrial symbiosis ontology with four levels of abstraction: i) meta-level, ii) top level, iii) domain level and iv) application specific level.

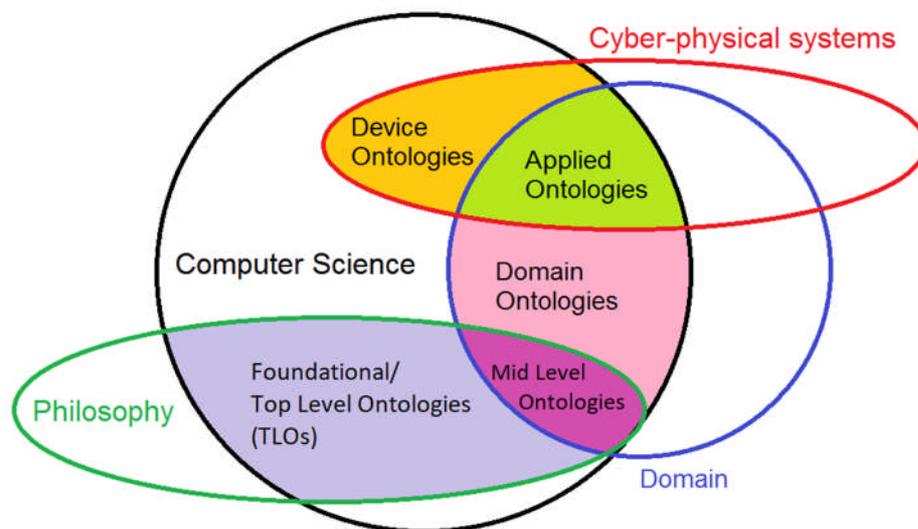


Figure 2. Venn diagram of infrastructure ontologies

The overlaps between ontology levels and disciplinary areas, including philosophy, cyber-physical systems, and computer science (encompassing knowledge engineering, data science and information management) are illustrated in Figure 2. The main focus of this report is on top level ontologies. Notably these are domain-independent. Ontologies used exclusively in philosophy are outside our scope, as are matters in computer science, cyber physical systems and domains which are unrelated to ontologies.

Information sub-sets found in domain ontologies are also to be found in both mid level ontologies and applied ontologies. Information sub-sets found in device ontologies are also found in applied ontologies. The only applied ontologies of interest here are those with interest in cyber-physical systems, since these are the ones required for digital twins.

In order to build up to the significance of top level ontologies, ontology levels are presented from the most specific to the most generalisable.

3.3.3 Applied ontologies

Applied ontologies are pertinent only to a specific purpose for which they are designed, considering just the objects and relationships that must be defined for that application. These are the most specific types of ontology, and may even be designed for a particular dataset, or include certain industry codes or standards. Application ontologies have been developed for purposes including traffic congestion forecasting (Prathilothamai, Marilakshmi, Majeed, & Viswanathan, 2016), flood risk analysis (Wu, Shen, Wang, & Wu, 2020) and building energy management (Marinakos & Doukas, 2018). Datasets relating to two or more applied ontologies cannot be integrated robustly without consideration of the top level ontology relevant for each applied ontology and creating a mechanism for interpreting one to another.

Stand-alone 'applied ontologies' meet specific use cases or specific purposes. Where datasets from two or more different applied ontologies need integrating, a mechanism to translate between their top level ontologies is needed.

3.3.4 Domain ontologies

Domain ontologies capture the concepts and relationships of interest within a particular domain. These ontologies bring together terminology specific to a single domain to address the needs of particular groups of users (Rudnicki, Smith, Malyuta, & Mandrick, 2016). Although domain ontologies seek to define a shared conceptualisation of a given domain, the unique approaches and perceptions of their individual designers means a single domain can be represented from different perspectives by numerous ontologies (Cummings & Stacey, 2017). For example, an ontology of the water domain could consider water systems through a technological, political, environmental, social, or economic lens, each of which captures only part of a complex network.

3.3.5 Device ontologies

Device ontologies deal with the representation of components that collect or distribute data within a system. This can include the observations of sensors, systems, and actuators, as well as their position, communication abilities, and manufacturing details (Szilagyi & Wira, 2016). Device ontologies have been described in recent work as 'Internet of Things' (IoT) ontologies,

as they are primarily concerned with sensor data, the backbone of the IoT. As they can facilitate the integration of multi-modal sensor data from different sources, device ontologies are typically not domain-specific (Bajaj, Agarwal, Singh, Georgantas, & Issarny, 2017). The most popular example of a device ontology is the World Wide Web Consortium's (W3C) Semantic Sensor Network (SSN) ontology (Compton et al., 2012).

3.3.6 Mid level ontologies

Mid level ontologies (MLOs) act as a bridge between TLOs and more specialised domain ontologies. They allow the high-level abstractions of the former and low-level details of the latter to be incorporated into a common semantic architecture (Rudnicki et al., 2016). MLOs are typically designed as an extension of existing TLOs, and thus usually developed using a top-down approach. While MLOs inherit the basic class hierarchy of their TLO counterpart, they also introduce domain-spanning knowledge and terminology. Examples of MLOs include the Mid Level Ontology (MILO) from the creators of SUMO (Niles & Terry, 2004) and the Common Core Ontologies (CCO) (Rudnicki, 2019).

3.3.7 Top level ontologies (TLOs)

Top level ontologies (TLOs), sometimes referred to as foundational ontologies, deal with the most abstract and fundamental ontological concepts. A foundational ontology can be defined as an ontology that “*defines a range of top level domain-independent ontological categories, which form a general foundation for more elaborated domain-specific ontologies*” (Giancarlo; Guizzardi & Wagner, 2004).

It is in TLOs that the philosophical elements of ontologies are most evident. TLOs contain generic terms, non-specific to any particular domain. They define high-level concepts, such as objects and events, and types of relationships between them, including parthood, participation, and dependence (Schmidt, 2020). Creators of TLOs have to make key ontological decisions, such as whether to adopt a perdurant or endurant stance, and how to prioritise past, present, future events. Examples of TLOs include:

- Basic Formal Ontology (BFO) (Arp, Smith, & Spear, 2015),
- Business Objects Reference Ontology (BORO) (de Cesare & Partridge, 2016)
- Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) (Gangemi, Guarino, Masolo, Oltramari, & Schneider, 2002).
- Suggested Upper Merged Ontology (SUMO) (Pease, Niles, & Li, 2002),
- Unified Foundation Ontology (UFO) (Giancarlo Guizzardi, Wagner, Almeida, & Guizzardi, 2015)

3.4 Domain (sectoral) ontologies

3.4.1 Infrastructure ontologies

Given various institutional, regulatory and organizational divisions, it is not surprising that infrastructure knowledge is distributed among various disciplines and sectors. Sectoral ontologies are described in the following paragraphs.

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” (El-Gohary & El-Diraby, 2010, p730). The Infrastructure and Construction PROcess Ontology IC-PRO-Onto aims to serve as a basis for “developing further model extensions, domain or application ontologies, software systems, and/or semantic web tools.” (ibid).

3.4.2 Energy sector ontologies

With numerous companies involved in the supply and distribution of energy, the use of ontology to provide a shared knowledge base is particularly valuable. As a result, of the main infrastructure sectors, it is in energy that ontologies have been most widely integrated. Energy sector ontologies found in academic literature are presented in Figure 3.

With data coming from such a range of sources, energy ontologies have utilised existing device ontologies to bring information together in a common format. Several energy ontologies have made use of the SSN ontology (Compton et al., 2012), a popular device ontology, to enable the integration of data from various sources (Corry, Pauwels, Hu, Keane, & O’Donnell, 2015; Dey, Jaiswal, Dasgupta, & Mukherjee, 2015).

Scale is another important consideration in the energy domain. While many applied ontologies focusing on household or building-level energy consumption, more ambitious domain ontologies have attempted to cover entire cities, districts or urban areas. Perhaps the most extensive example is the EU’s SEMANCO project, which produced an urban energy ontology as part of its objective 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 (Madrado, Sicilia, & Gamboa, 2012). The work of SEMANCO lives on in ‘Energy Efficient Cities’ (EECities), a new project to support planners and policy makers using technologies based on the SEMANCO platform.

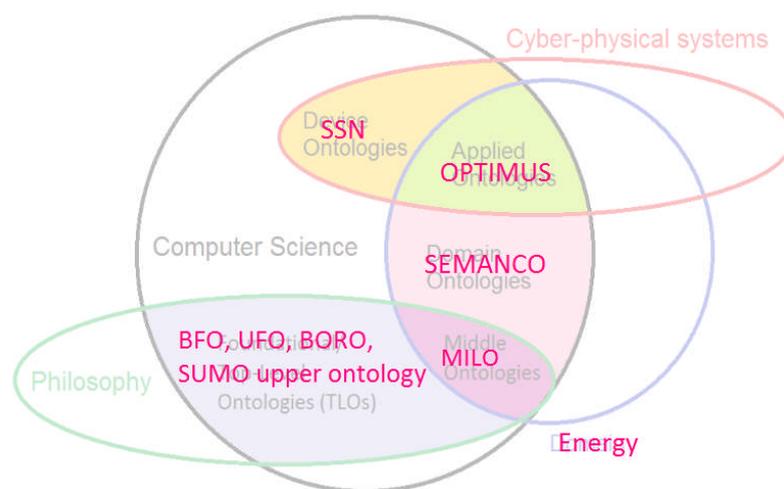


Figure 3. Venn diagram of infrastructure ontologies overlaid with ontologies from energy sector

While energy ontologies exist at both the domain and applied level, there are also examples which incorporate existing elements of higher-level ontologies, such as the OPTIMUS energy ontology, which utilises classes from both the SEMANCO and SSN ontologies, for the purpose of smart energy management in buildings (Marinakakis & Doukas, 2018). With the SEMANCO ontology itself designed to align with the SUMO TLO, this is an example of integration across a broad spectrum of abstraction, from defining classes to an applied use case.

3.4.3 Water sector 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.

For these reasons, there have been few domain ontologies developed for the water domain. Perhaps the broadest ontology attempted in this sector is the water supply ontology that underpins 'WatERP', an open online platform which aims to coordinate the management of supply and demand in order to reduce water usage and associated energy consumption (Varas, 2013). Water sector ontologies found in academic literature are presented in Figure 4.

Many of the attempts to date to create ontologies for water infrastructure can be considered applied ontologies, considering only sections of the water domain pertinent to their application. These can prove very valuable in tackling real-world challenges. For example, an ontology-based approach to disaster risk evaluation has allowed researchers to identify the key influences behind urban flooding in Zhengzhou City and make suggestions for managing future threats (Wu et al., 2020). Ontologies can also assist in automated decision support for identifying and mitigating failures in the water distribution network, helping to maintain a dependable water supply (Lin, Sedigh, & Hurson, 2012).

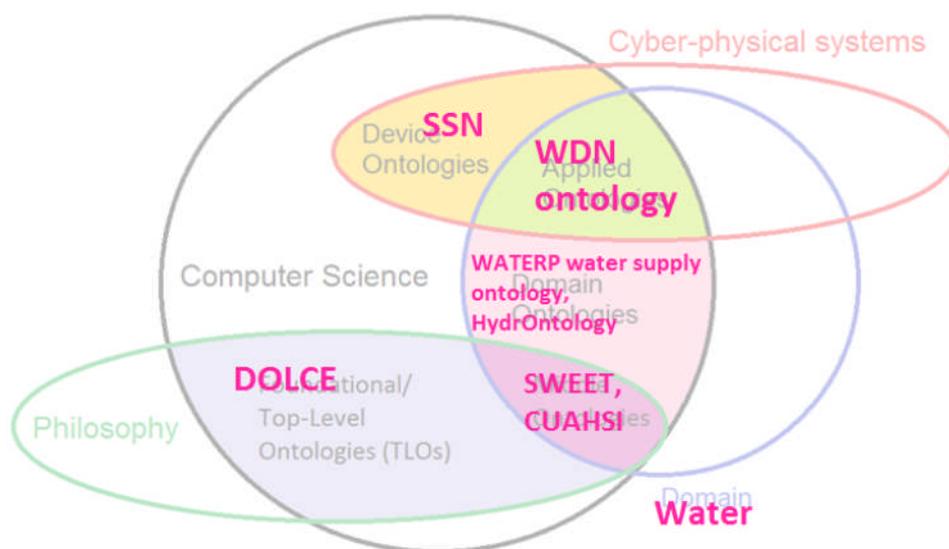


Figure 4. Venn diagram of infrastructure ontologies overlaid with ontologies from water sector

Although far from commonplace, there are examples of integration of different levels of ontology in the water sector. As in energy management, there is a need to collect data, such as pollutant levels, using sensors. 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 MLO SWEET (Semantic Web for Earth and Environmental Terminology). In the important application of water quality management, elements of these existing ontologies have been incorporated into an applied ontology which combine sensor readings with regulatory and pollutant data, to ensure the water we consume is safe (Ahmedi, Ahmedi, & Jajaga, 2013).

3.4.4 Transport sector ontologies

Perhaps unusually, when compared to other infrastructure domains, the transport sector has seen numerous attempts at domain ontologies, albeit varying in scope. More in line with other sectors, applied transportation ontologies have also been developed, often specific to an application for a particular mode of transport. Transport ontologies found in academic literature are presented in Figure 5.

In the transport sector, the boundaries for what constitutes a network are much clearer than, for instance, the water domain. This could explain why there have been significantly more attempts at domain ontologies in transport than in other infrastructure sectors. 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 line with what is seen in other infrastructure sectors, there are numerous applied ontologies that demonstrate the problem-solving value of ontology usage in the transport sector. It can be noted that a significant proportion of applied ontologies in transport are based around road usage. Ontologies have been developed to manage and reduce congestion on public roads (Abberley, Gould, Crockett, & Cheng, 2017; Prathilothamai et al., 2016), benefitting both drivers and the environment. Another focus of applied transport ontologies has been road accident identification and response. Ontologies designed for this purpose typically centralise decision-making, relying on in-vehicle sensors to provide data (Barrachina et al., 2012).

While ontologies are typically utilised by those managing transport networks, applied ontologies have also been incorporated into systems designed to be operated by individual users. One such example proposes a way for user to interact with a public transport ontology in order to effectively plan their journey (Mnasser, Oliveira, Khemaja, & Abed, 2010), while another combines ontology with natural language processing to enable drivers to query a Twitter-based system to receive relevant traffic information (Wanichayapong, Pattara-Atikom, & Peachavanish, 2015) An alternative approach to accident response also takes an individual approach, integrating an ontology into each vehicle to assist nearby drivers in effectively reacting to emergency situation (Bermejo et al., 2014).

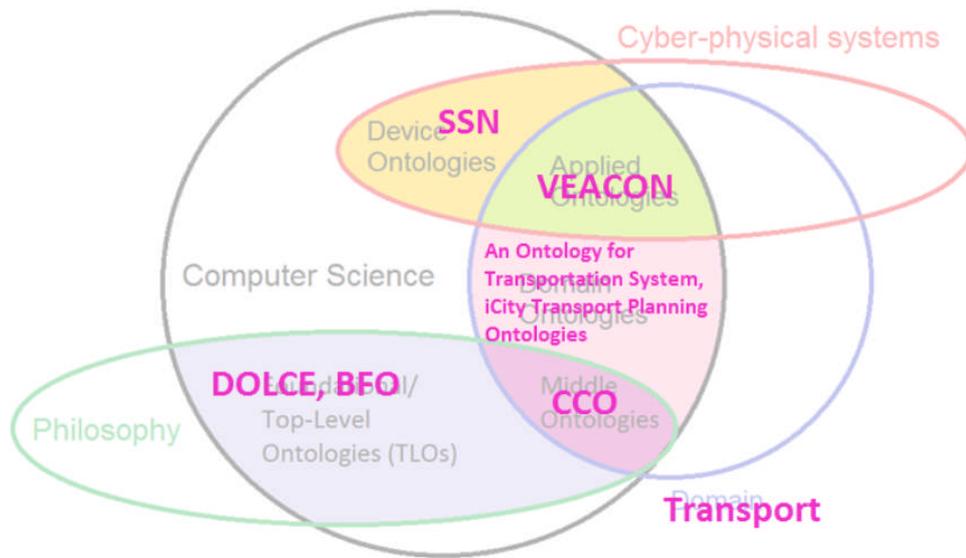


Figure 5. Venn diagram of infrastructure ontologies overlaid with ontologies from transport sector

3.4.5 Telecoms sector 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. However, the diversity of technologies and applications at the edges of digital infrastructure can lend themselves well to application ontologies, while a handful of works have taken this a step further, proposing ontologies with both domain and application levels.

Interestingly, the telecoms sector has also seen domain-specific ontological languages proposed, although these predate 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). The later comprises of both a broader domain ontology, and an application ontology for the input and output of services.

Some application ontologies in telecoms have focused on specific types of network. One example has sought to simplify the configuration of 3G wireless networks (Cleary, Danev, & Donoghue, 2005), while another has been proposed for optical transport networks based on the ITU-T G.805 and G.872 recommendations (Barcelos, Monteiro, Simões, Garcia, & Segatto, 2009). Others have concentrated more broadly on mobile ontologies. As part of the SPICE project, Villalonga et al. (2009) attempted to standardise ontologies in this sector, developing a core ontology with sub-ontologies for services, user profiles, content: presence, context, and communication resources. Uzun and Küpper (2012) instead take a 'Linked Data' approach, incorporating data describing mobile networks, their topology and components (e.g., base stations, mobile devices and Wi-Fi access points) into their OpenMobileNetwork.

While application ontologies may be more prevalent than domain ontologies in the field of telecoms, a handful of ambitious works have developed ontologies that are broader in scope.

The authors of Telecommunications Service Domain Ontology (TSDO), suggest a modeling approach that brings together six sub-ontologies - the Terminal Capability Ontology, Network Ontology, Service Role Ontology, Charging Ontology, Service Quality Ontology, and Service Category Ontology – to address the challenge of semantic interoperability (Qiao, Li, & Chen, 2012).

Perhaps one of the greatest difficulties in representing systems in this sector is the increasingly complex and heterogenous nature of telecommunication networks. 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 (Zhou, 2018). Recent work has expanded upon this, demonstrating how, by describing the core concepts of devices, interfaces, links, data and service, ToCo can be applied to hybrid telecommunication networks (Zhou, Gray, & McLaughlin, 2019).

It is this view of networks as systems of devices that may explain the adoption of device ontologies within the telecoms domain, where such ontologies are more commonly reference to as 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 base ontology is specifically designed for interoperability for IOT (OneM2M, 2021) and is built into 4G in the SCEF function. These are likely to be essential in making Smart Contracts work (Digital Catapult, personal communication).

3.4.6 Sewage and solid waste sector 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, has been developed by the academics behind the Sm@rtBin project in Malaysia. The ontology, which aligns with the UFO TLO, has been evaluated through its performance in applying the Sm@rtBin prototype, a QR code-based centralised system for monitoring the collection of waste bins and dumpsters (Ahmad, Badr, Salwana, Zakaria, & Tahar, 2018). This is a fairly unique example of a domain ontology which utilises TLO architecture yet is validated through an application-specific case study. Demonstrating through task-based use cases how the OntoWM ontology can effectively capture and store the knowledge relevant to the sub-domain of waste collection, its developers seek to prove that OntoWM, and indeed ontologies in general, can benefit the broader domain of waste management (Sattar, Ahmad, Surin, & Mahmood, 2021).

Waste management can go beyond simply collection and disposal. 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) have created an ontology to represent the domain of

industrial symbiosis. Industrial Symbiosis (IS) is a growingly accepted paradigm for processing waste into material, energy and water. In their ontology, they treat waste as a resource, and use semantic matching to distinguish the difference between classes to create an IS network. This ontology has been integrated into a web-based platform, which has been in operation for several years, facilitating IS in numerous companies across Viotia, Greece (Cecelja et al., 2015).

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. In an interesting example, Recycleye have compiled a database of over 2.5 million images of waste, as part of their goal to integrate computer vision and robotics into waste management (Recycleye, 2020). Dposal, 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 (Dposal, 2021, Onerhime, 2020).

3.5 Infrastructure research 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.

3.5.1 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 enduring 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.

3.5.2 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.

3.5.3 Ontological stats 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.

3.5.4 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. See Figure 6.

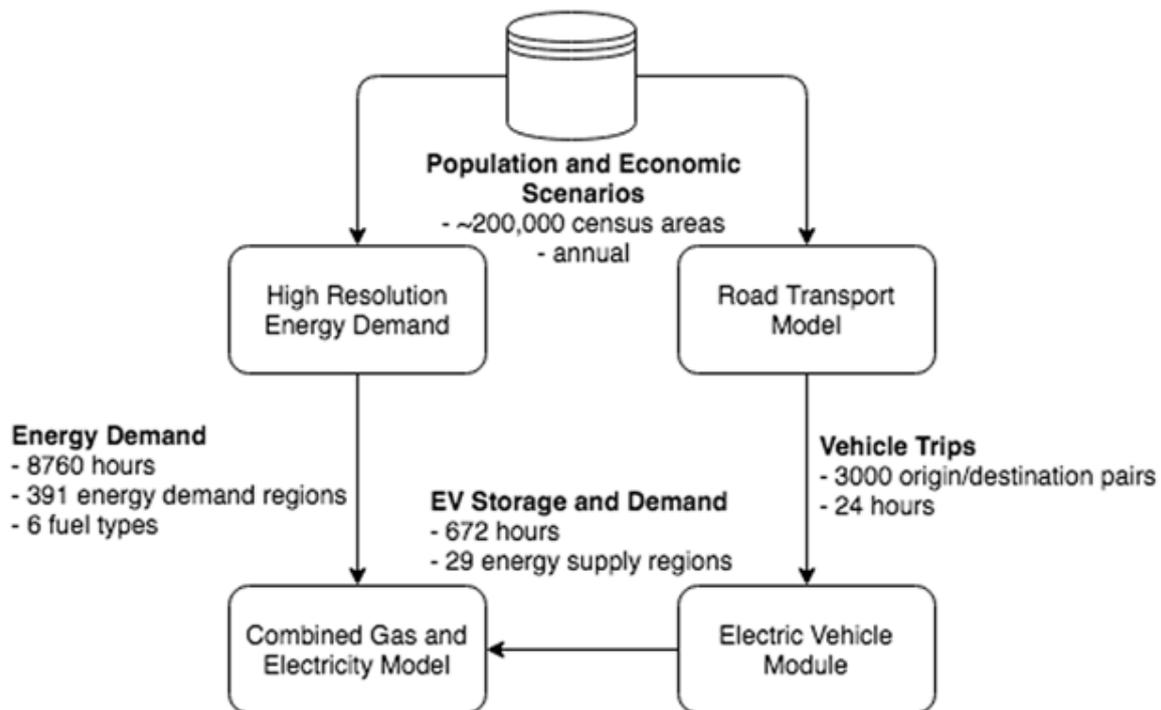


Figure 6. Diverse data dimensions across infrastructure. Usher & Russell (2019).

3.5.5 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 min/max 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.

Ontologies are needed to explicitly represent definitions of extent and granularity, noting that definitions change over time, and that there is no single preferred scale.

3.6 Ontologies of ontologies

3.6.1 Infrastructure as a complex system

Infrastructure systems are complex and do not exist in isolation. It is therefore essential that infrastructure is described (with reference to ontologies) in terms of its relationship to other systems, processes, people, and procedures. For example, the maintenance regime for a structure, such as a bridge, is a function of its design, the outcome of computational models, material and method of construction, previous observations and repairs, its use and loading, and applicable standards and regulatory frameworks. This example spans many professions and disciplines and should not be considered by a single ontology.

An approach based on ontologies of ontologies is required, in which there is sufficient commonality between the overlapping regions of each ontology that they can be usefully combined without excessive ambiguity. Physical infrastructure itself can be related to its surroundings by its geography, hence this section considers integration mainly from a spatial data perspective.

3.6.2 Semantic interoperability

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 sector 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.

The standards that underpin BIM are mature and in widespread use, partly due to the BIM mandate for public sector projects in the United Kingdom. Interoperability beyond the level of individual buildings and plots however becomes more complex: geospatial standards such as those from OGC, CityGML and LandInfra have significant differences to IFC but also substantial areas of overlap (Vilgertshofer, Amann, Willenborg, Borrmann, & Kolbe, 2017, Gilbert, Rönsdorf et al., 2021).

The extent to which data in a BIM for a project survives movement from a capital works project to facilities and operational management is a cause for concern. Anecdotally that seems to be a significant point of failure yet digital twins will need access to this data particularly for use cases that are about discovering instances of particular materials, methods, standards, etc.

3.6.3 Scale and algorithmic reconciliation

Even when the semantics can be aligned, there is often a data gap that exists between the building model and the wider utility and infrastructure networks that requires detailed consideration or complex algorithms to resolve (Gilbert, James, Smith, Barr, & Morley, 2021). This remains an area of considerable research effort, focusing on tackling the differences in data models, geometry representations, scales, uses and purpose, and coordinate systems (Noardo et al., 2019). Vilgertshofer et al. (2017) demonstrate a methodology whereby building and city data is combined using ontologies and using semantic web technologies (Resource Description Framework - RDF) but highlights a dependency on shared identifiers in each dataset.

The INSPIRE directive is the most notable example of an international (pan-European) standard that seeks to establish a spatial data infrastructure across multiple themes. Whilst the underlying data models for this were not originally defined as ontologies, there are numerous examples of successful conversions (e.g. Tschirner, Scherp, & Staab, 2011) and an official draft ontology now exists for many of the spatial object types (Echterhoff & Portele, 2017).

Information such as the topology of the road network can be represented using the INSPIRE ontologies but can't be represented in CityGML without using an application domain extension (ADE) (Beil, Ruhdorfer, Coduro, & Kolbe, 2020). This flexibility in CityGML makes the combination of INSPIRE and CityGML a strong candidate for an overarching ontological framework used to integrate other, more domain specific ontologies and data.

Considering the Grenfell example from the introduction of identifying properties with potentially hazardous cladding, it is conceivable that a combination of the INSPIRE ontology for cadastral parcels, CityGML and BIM data for construction, management and risk ontologies (e.g. El-Gohary & El-Diraby, 2010) and material properties (e.g. Ashino, 2010) could have enabled a rapid assessment for all buildings. This could be further extended by incorporating models, such as for slope stability, which would enable a digital twin approach for looking at landslip risk following extreme rainfall.

The lack of a widely adopted overarching ontology for the primitive and key concepts for infrastructure systems currently prohibits various use cases.

3.7 Ontology languages and Resource Descriptors

3.7.1 Resource Description Framework (RDF)

The Resource Description Framework (RDF) is a framework for representing information in the Web. RDF enables structured metadata to be encoded, exchanged and reused due to the support provided by RDF for consistent encoding and exchange (E. Miller, 1998).

RDF, developed by the W3C, uses Universal Resource Identifiers (URIs) to define a subject, object, and relationship between them. These can be called 'triples'. More specifically, language is made of sentences called 'triples' because they have three elements: a subject, a predicate and an object. The subject is a resource (hence identified by a URI). The predicate is in effect a relation between the subject and the object. The object may be a resource or a data type value.

3.7.2 Ontology Web Language (OWL)

Ontology languages have been proposed for the Semantic Web, to give formal semantics for ontology creation. In 2004, the Ontology Web Language (OWL) was presented by the W3C, and soon became the industry standard. A revised version, OWL2, followed in 2009 (Grau et al., 2008). Almost all modern formal ontologies in the fields of computer science, engineering, medicine and others, are now written in OWL2.

3.7.3 Semantic extensions

OWL can be seen as a semantic extensions of RDF, utilising the RDF syntax for RDF Schema RDF(S). RDF(S) can be mapped into a syntax suitable for OWL2, however this second iteration of OWL was also extended to include XML syntax as an alternative (Grau et al., 2008). OWL2 ontologies are primarily exchanged as RDF documents. Thus RDF(S) or OWL have XML serialisations. An ontology differs from an XML schema in that it is a knowledge representation,

OWL2 ontologies are defined in terms of classes, properties, and individuals, which are members of classes. The semantics of a class is defined by its relationship to other classes, through concepts such as subclass, conjunction, disjunction, and negation. This allows conditions to be defined under which a given individual can be a member of a particular class (Katsumi & Fox, 2018).

3.8 Literature on tooling, technologies, and methods

3.8.1 Technological options

Ontologies represent only one component of the larger picture of integrating disparate systems and models for digital twins. The ontologies provide commonality to the understanding and vocabulary through which concepts are expressed, but commonality may also need to exist in the interaction patterns – how one system communicates with another or answers queries – and the tooling and technologies that enable and reduce the complexity of achieving interoperability. This section explores some of the technology options available for ensuring infrastructure ontologies become findable, accessible, interoperable, and reusable in line with

the FAIR principles (Wilkinson et al., 2016). It is therefore predicated on the assumption that this approach will have much in common with linked data technologies on the web, even if sometimes the data resides in standalone systems or is not made publicly available.

3.8.2 Application programming interfaces (API)

An application programming interface (API) is the technical means of defining interactions between software. These interfaces can be and are increasingly web-based, meaning requests and responses follow many of the same principles as accessing and interacting with web pages (i.e. communication takes place using HTTP using its standard verbs). Interoperability is thereby achieved using a common understanding of the protocol (HTTP), the concepts (ontologies and vocabularies), machine-readable data (involving schema, formats, and serialisation), and an understanding of what options are available and how to formulate a request for one of those options (a predefined set of options, part of hypermedia) known as the HATEOAS principle per Fielding (2000).

APIs that meet these expectations are said to follow the de-facto standard of representational state transfer (REST) as described by Fielding (2000). In theory, any person or system should be able to interact with a REST API without any knowledge other than understanding the protocol and format. This is analogous to any web browser being able to access any website because the browser understands HTML and is able to render it on screen but differs in that the human rather than a machine interprets the web page, considers its contents, and makes decisions about which links to follow.

These advances and design patterns that now underpin a huge number of online services also provide means for expressing infrastructure research data, defining its relationship to ontologies, and making it available to third parties. Recent research has demonstrated that in some cases data expressed in ontologies can be used to automatically generate REST APIs (e.g. Garijo & Osorio, 2020), and some widely used REST APIs can be related back to ontologies (e.g. Toghiani & Kameas, 2012). In making the jump from infrastructure data to infrastructure ontologies and a technical framework that supports digital twins, APIs are likely to be the means through which different components of a digital twin interface are exposed. It is therefore necessary to explore API standards that facilitate linked data.

3.8.3 Serialisation

The need to combine ontologies with web technologies, alongside efforts to make more of the web machine readable, have heavily influenced the way in which we map data to ontologies. The fundamentals of RDF and triples can be expressed in many forms including Turtle (Prud'hommeaux & Carothers, 2014), N-Triples and N-Quads (Carothers, 2014) and JSON-LD (Kellogg, Champin, & Longley, 2020). RDFa (Adida, Birbeck, McCarron, & Herman, 2015) and Microdata (Neville & Brickley, 2021) also provide a way of embedding these ontology references within web pages and similar structured documents as properties of other data.

Roughly three-quarters of web-based APIs already use JavaScript Object Notation (JSON) as their serialisation format. JSON-LD is a fully compatible extension to JSON for linked data purposes and integrating with ontologies and is now recommended by major search engines, aiding adoption. There are numerous other examples where embedding ontologies within existing formats have led to far greater uptake (Lanthaler & Gütl, 2012). Note also schema.org and schema representation within unstructured, semi structured, and soupy HTML markup via microformats.

It is possible to convert JSON-LD representations to N-Quads and other linked data formats relatively easily, and back again. This conversion should be lossless though not always identical, as the same data can be expressed in different but equivalent forms using JSON-LD (Kellogg et al., 2020).

Non-web-based data concerning the built environment such as BIM is unlikely to be serialised as JSON based on present standards and software in use, with IFC-SPF being the most widely used format today. However, BIM data expressed in ifcXML (Nisbet & Liebich, 2007) could be extended using RDFa or Microdata (Hor, 2015). This could be useful alternative to schema mapping (Gilbert, Rönsdorf, et al., 2021). It is also possible to express BIM data in RDF form, aligned to ifcOWL. Both of these options would provide opportunities for integrating infrastructure data with ontologies other than BIM and for greater interoperability with other software.

3.8.4 Validation and suitability

Both ontology verification (ensuring ontology definitions implement requirements) and validation (ensuring the ontology models the real world) are important considerations in ensuring their suitability and uptake in a given domain (Gómez-Pérez, 2004).

A range of quality criteria have been identified to support ontology verification, including accuracy, adaptability, clarity, completeness, computational efficiency, conciseness, consistency or coherence, and organisational fitness (Vrandečić, 2009). It is important to note that a good ontology may not perform equally well against all these criteria, with some being in opposition to each other. Evaluation can take place against different aspects of an ontology, including its vocabulary, syntax, structure, semantics, representation, and context (Vrandečić, 2009).

Validation can be considered in terms of conceptual validity; constraints, meaning whether the data matches the constraints and expectations laid out in the ontology; and application, meaning whether the data is suitable for a specific use case. Automated approaches for validation have been available for some time now, for all aspects apart from the design and structure of an ontology, for which consensus and in-depth discussions serve a valuable purpose, but these tools are not widely used in practice (Brank, Grobelnik, & Mladenić, 2005). There is evidence that many users of linked open data and Microdata are making errors such as using undefined properties or using them in the wrong ranges (Paulheim, 2015).

Recent developments in web standards may assist in making validation tools more accessible and encouraging adoption. For example, the OpenAPI standard is recommended by the UK Government (GDS, 2020) and was recently aligned with JSON Schema. The JSON Schema standard (Wright, Andrews, & Button, 2021) provides a means of validation that isn't as flexible or comprehensive as an ontology defined in OWL, which can be challenging to validate using because of its open world assumption (Sirin, 2010) but is implemented in software libraries (e.g. ajv) and simple enough to be used in the completion of web forms such as when entering metadata.

Non-compliance with published ontologies and schemas is by no means a new problem. In some cases, ontologies are even hijacked for another purpose because no suitable ontology was clearly defined (Hogan, Harth, Passant, Decker, & Polleres, 2010).

3.8.5 Accessing, querying and interacting with data and its ontologies

In a survey of existing tools for accessing, querying and interacting with linked data, Klímek, Škoda, & Nečaský (2019) defined 36 requirements and 93 evaluation criteria for a hypothetical Linked Data Consumption Platform. Core functionality including enabling users to find data; summarizing the data of candidate datasets based on the (meta)data of that dataset; recommending related datasets; offering data transformation between different representations; choosing entities and properties that are required for a given goal, and preparing data for visualisation. However, no existing tools were identified as supporting all aspects of the linked data consumption process, posing limiting the practical adoption of linked data by stakeholders.

Mechanisms for interacting with linked data depend on the purpose. Queries akin to those used with databases can be achieved using SPARQL (Prud'hommeaux & Seaborne, 2008), an extension of SQL that is used by almost all database engines. A limitation of SPARQL however is that while queries can be federated across disparate datasets, the computational burden rests with the data provider (on their servers). Verborgh et al (2016) present an alternative approach of 'linked data fragments' that bridges an API approach with SPARQL with reduced server load. Not all data publishers are going to be willing or able to make the entirety of their data available and may instead elect to support predefined queries or expect third parties to use their own model process rather than developing their own. These interaction patterns (as opposed to data sharing) can be handled using the REST API approach, with the interaction patterns themselves described using ontologies and vocabularies such as JSON Hyper-Schema (Andrews & Wright, 2019). The complexity of writing a generic client for these APIs means there are very few implementations.

3.8.6 Visualisation

Adoption of linked data beyond the semantic web community has been limited, with challenges including the large amount and heterogeneous nature of linked data (in terms of granularity, completeness, its use of ontologies, source- and target-domains) (Dadzie & Pietriga, 2017). There exists a significant knowledge gap between semantic web experts and lay users in other domains. Ontologies provide a structure and a practical way of categorising and retrieving data and can be used to aid interpretation of meaning across multiple domains, however the use of ontologies to structure linked data is restricted to technical experts. While visualisations of linked data can assist with these challenges, they are not a panacea as they have to be tailored to specific tasks and support diverse users at the relevant level of detail and abstraction. Key tasks for visualisation as identified by a 2017 review (Dadzie & Pietriga, 2017) include the identification of entry points to linked data, navigation within and across multiple linked datasets to support exploratory discovery, the analysis of data structures and their alignment, querying, sense-making and guided in-depth analysis, content enrichment (annotation and the identification and derivation of new links), and the presentation and sharing of data and results.

3.8.7 A supporting ecosystem

An ecosystem is required, in addition to the technical capability, whereby ontologies are updated to meet developing needs (which means abuse of an ontology is something to identify and respond to), and software libraries, tools and schema are widespread and work to reduce complexity for users. As an example, the content management system Drupal adopted

schema.org leading to adoption by many users who are unaware they are working with the ontology (Paulheim, 2015).

Publishers of data and ontologies should also be consumers. This concept, often referred to as dogfooding (Harrison, 2006) can help to identify potential issues, ensure that the data works for various use cases, and can help to avoid a two-tier approach to data publishing where only a delayed or reduced subset is made available for general use.

4 Industrial Practices

This section presents results of reviews on industrial practices and from practitioner engagement.

The first sub-section describes the Data Analytics Facility for National Infrastructure hosts datasets, models, and visualisations used in infrastructure research. These components may be integrated by a workflow process enabled by standards, methods and tools to classify, and organise the information.

Next, industrial ontologies are identified and described, followed by mid level ontologies which are having a role in integrated environments such as city models and digital twins.

Data is also collected on barriers, risks, opportunities and strategies, from practitioners in energy, transport and water systems: similarities and idiosyncrasies are identified. To end the section, the results of a poll on governance, trialling, and addressing capability gaps are presented in the context of digital twins.

4.1 Data and Analytics Facility for National Infrastructure (DAFNI) data

4.1.1 DAFNI purpose

The DAFNI platform (STFC, 2021) offers a powerful research infrastructure service able to receive and operate simulation models and visualise the results, as well as to hold and make available many hundreds of substantial 'legacy' data volumes, concerning themes such as built infrastructure networks, environmental, socio-demographic profiles and cadastral (land) information.

Instrumented and archived data need to be held together in ways that allow its decoupling from the observed systems, permitting scenario modeling of various management strategies and exploration and visualisation of the potential environmental consequences. The DAFNI facility offers a platform that can provide this, and it will be possible to ingest real-time sensor data to be placed alongside extant static data volumes, with all these being made directly available to suites of modeling and visualisation tools.

4.1.2 Top level categories and ontologies

Substantive dataset resources are included within DAFNI, and an approach is required to facilitate data discovery, and linking of key datasets in modeling applications. Consideration must be given to the potential for ontologies to serve and improve the data classification and search tools (taxonomies), and to develop and document approaches for metadata descriptions for data holdings.

DAFNI adopts the use of the INSPIRE Topic Categories as data descriptions both to classify newly added data items and for enabling efficient searching of existing data. The INSPIRE Directive (European Commission, 2007) aims to “create a European Union spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment. This European Spatial Data Infrastructure will enable the sharing of environmental spatial information among

public sector organisations, facilitate public access to spatial information across Europe and assist in policy making across boundaries.”

Additional taxonomies such as the European Science Vocabulary (EUSciVoc) (CORDIS, 2020) could be incorporated into DAFNI alongside the INSPIRE Topic Categories to increase the granularity and adopt a hierarchical approach to aid data discovery.

4.1.3 Standards alignment

To protect the development of the DAFNI platform (and for future proofing) international standards have been adopted from trusted institutions like **ISO** and **W3C**. Developments from public institutions at all levels of government, as the main sources and eventual destination for datasets, were studied before adoption of standards. To help with future interoperation, there are numerous alignments with e.g., the **ISO 19115** standard and **INSPIRE** guidelines (European Commission, 2007), and the UK version **GEMINI** (AGI, 2015). This may help users to comply with the latter as compliance is a legal requirement for geospatial datasets from public bodies (DEFRA, 2015). See also data.gov.uk (2021).

4.1.4 Interoperability

The DCAT2 metadata standard (W3C, 2020) adopted in DAFNI provides “an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web”, offering a structured and formalised means to characterise the datasets held in DAFNI. DCAT-2 is able to describe datasets, resources, services and catalogues. DCAT2 further permits datasets to be classified with wider thematic classifications. For this, it uses the Simple Knowledge Organization System (**SKOS**) (SWD WG, 2012) a common data model for sharing and linking knowledge organization systems via the Web.

Using SKOS, concepts can be “identified using URIs (Universal Resource Identifiers) labelled with lexical strings in one or more natural languages, assigned notations (lexical codes), documented with various types of note, linked to other concepts and organized into informal hierarchies and association networks, aggregated into concept schemes, grouped into labelled and/or ordered collections, and mapped to concepts in other schemes” (Miles & Bechhofer, 2009).

The USA and EU as well as national governments have officially adopted open semantic linked data to describe their data holdings from various departments also following FAIR data principles. The metadata system adopted in DAFNI should where possible be aligned to the adopted metadata systems in other comparable and associated data catalogues, to help ensure cross-compatible searching and other ‘FAIR’ data management characteristics.

DCAT and its associated namespaces have been the building block for their respective profiles, **DCAT-US** and **DCAT-AP** (pan-European). DCAT2 incorporates the amendments created by DCAT-AP, which had in turn extended the original version of DCAT. These will be further evolved into new profiles according to national bureaucratic procedures’ additions (national profiles). Moreover, key domain specialisations, like **GeoDCAT-AP** for geospatial datasets, are being widely adopted.

CERN and other scientific institutions also have adopted similar encodings to describe their datasets. Google and other search engines support **schema.org** as well as DCAT to encode their metadata. This has great implications for dataset discovery on the web. They all have in common that the serialisation format is **JSON-LD** which is both valid **RDF** and **JSON**.

DAFNI adopted a **DCAT2** (February 2020) based profile serialised in **JSON-LD 1.1** named **DAFNI-LD**. DAFNI-LD is highly extensible and it is offered to the users at large as a sandbox to experiment. The research community are able to make suggestions to shape DAFNI-LD. Accounting for unforeseen needs, all ontology levels could be reused in our metadata. The DAFNI platform is expected to selectively trial any such developments.

4.1.5 Dataset metadata highlights

The following are dataset metadata highlights:

- Good alignments with **Dublin Core** and **schema.org**
- Official contact point (note: not from DAFNI)
- Official licensing (also, further rights property as licence addendum)
- Permanent IDs: ORCID, ror.org, DOI, company registration number (CRN) or URI, etc.
- Standards compliance records
- Dataset provenance and attribution. (DCAT2 inc. **PROV-O**)
- Temporal information
- Geospatial dataset coverage (based on **geonames.org** but not limited to it)
- Geospatial dataset information in **GeoJSON-LD 1.0** – this is the basis of the larger **OGC** Earth Observation GeoJSON-LD standard and is compatible with RFC 7946 GeoJSON
- Beside keywords, the dataset classification is based on INSPIRE *topic categories* and its official *themes* vocabulary
- Rich file level metadata (distribution view)

4.2 Industrial Ontologies

Ontologies used in industry are strongly aligned with standard and recognised classification systems.

4.2.1 The Industrial Ontologies Foundry (IOF)

The Industrial Ontologies Foundry (IOF) (2021) initiative was to create a suite of ontologies to support digital manufacturing and promote interoperability in related fields. The IOF uses basic formal ontology (BFO) as its top level ontology. BFO is in the final stages of review to become international standard ISO/IEC PRF 21838-2.2. The IOF focus is on support of ontologies for design, maintenance, supply chain, production and lifecycle management in manufacturing. As a proof of concept for the BFO compliant ontologies, the IOF members proposed an open access, reference ontology (ROMAIN) for the maintenance management domain. Additionally, an IOF proof of concept paper was published by the IOF community that covered the developments within the project and provided information on the IOF organisational set up.

The notional levels of ontology, and the scope of the IOF ontologies, is developed in the IOF Charter (2021) as shown in Figure 7. The scope excludes application and domain dependent ontologies.

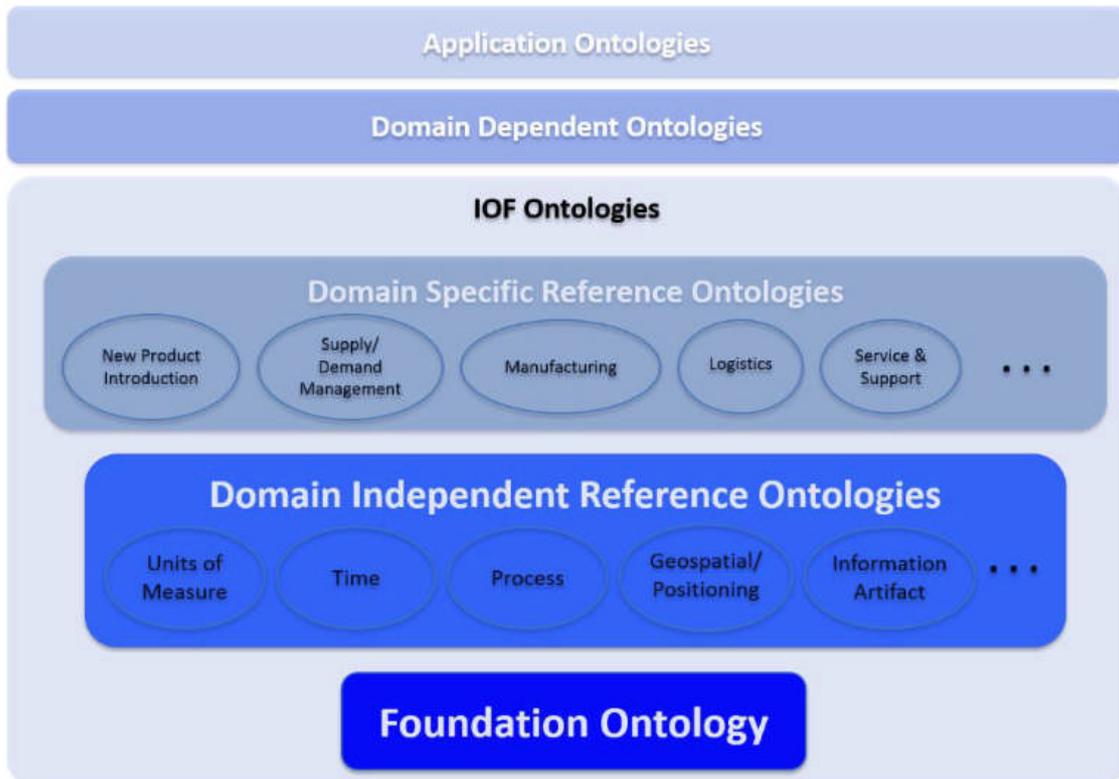


Figure 7. The architecture of the IOF ontologies (Industrial Ontologies Foundry, 2021)

4.2.2 Industrial data models and reference data libraries

The application areas of industry data models and reference data libraries identified in the Construction Innovation Hub survey (Leal, Cook, Partridge, Sullivan, & West, 2020) are shown in Figure 8. There are substantial overlaps between application areas, and for each model there are differences (Leal et al., 2020) in the:

1. Defining organization
2. Objectives and scope
3. Structure of the model
4. Extensibility
5. Documentation
6. Maintenance and usage

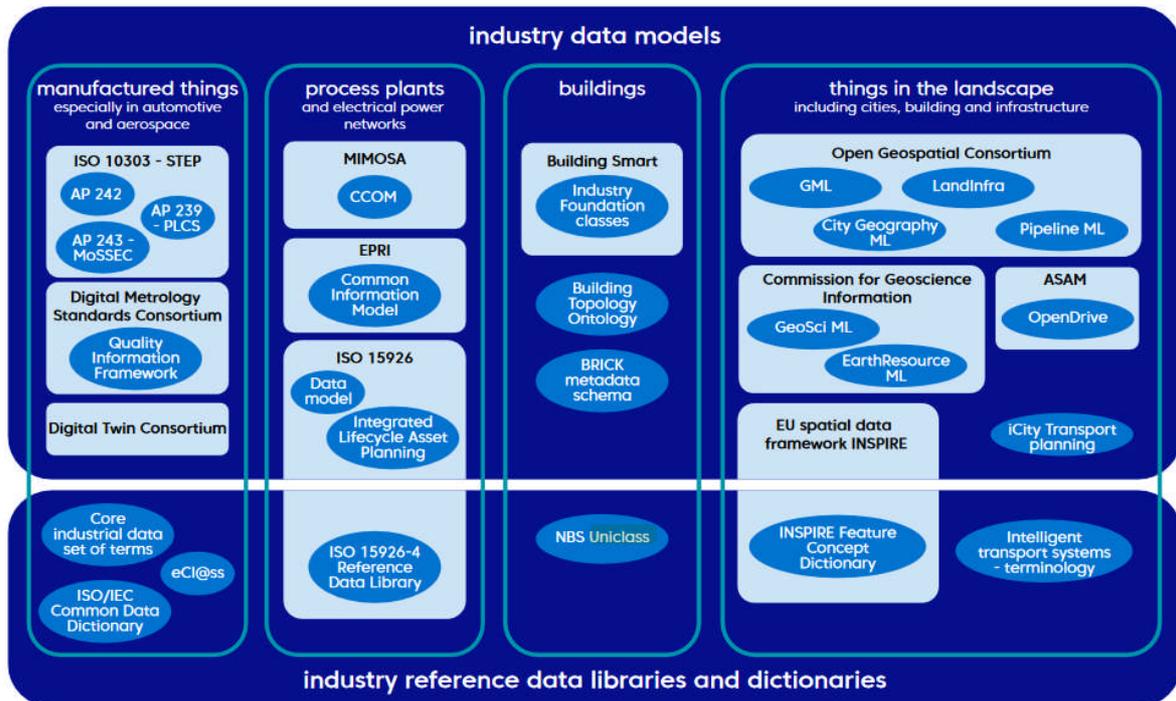


Figure 8. Industry data models and reference data libraries by application area (Leal et al., 2020)

4.2.3 Building Information Modeling (BIM)

Uniclass is widely used for BIM level 2 and compatible with some generic ISO standards and can be extended. It is an industry data model and is actively maintained by NBS. Industry Foundation Classes (IFC) provides a hierarchy of objects but not on a systems level. The (somewhat outdated) COBie used to do this via a dedicated list of all types.

Gao et al., (2015) define a process of conceptualizing and formalizing BIM knowledge from the IFC schema to construct a IFC IR domain ontology. The IFC IR Ontology can be used for the disambiguation of terms on online BIM documents. Whilst Lee et al., (2016) demonstrate an ontology-based approach for developing data exchange requirements and model views.

The UK BIM Alliance are currently preparing a 'position statement' on digital twins to clarify the relationship between BIM and Digital Twins, in alignment with discussions in various CDBB work groups. If the NDT is primarily a UK Government proposition to enable a greater understanding of the UK infrastructure then by its very nature, it will (most of the time) include built assets in one form or another. Digital Twins will hold dynamic data but the current 'typical' BIM has static data and is also often confused with 3D models. The distinction between the two is important but not necessarily that clear-cut as the BS EN ISO19650 process *can* (should?) also be used to procure a digital twin (Casey Rutland, personal communication).

4.2.4 Equinor (formerly Statoil) ontology (energy sector)

A petroleum company Equinor has carried out mapping of exploration data via Ontology Based Data Access (OBDA) to create an industrial ontology, aiming to abstract away from the schema level details of their data and conceptualise it in a clear manner (Kharlamov et al.,

2019). During this collaborative work a deployment module has been developed to create ontologies and mappings from relational databases in a semi-automatic fashion.

Due to the lack and nonexistence of documentation specifically related to the exploration work of Equinor geologists these ontologies were created. They were intended to eliminate complexity across different schema elements and mediate between data users and data sources. This is shown in Figure 9.

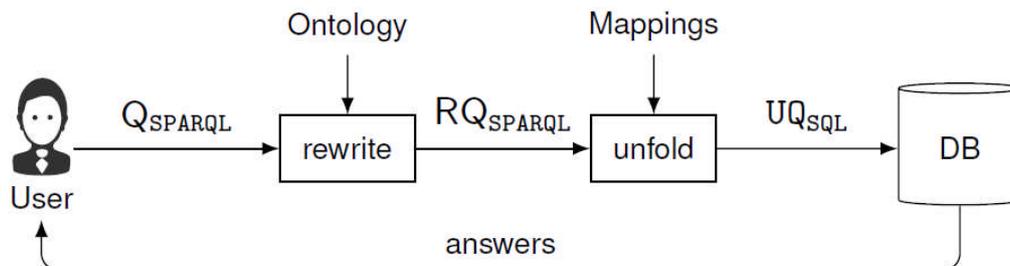


Figure 9. Equinor ontology based mediation between data users and data sources (Kharlamov et al., 2019)

The main goals were to create a single point of semantic data access and provide a user-oriented conceptual data model. The focus was on the two main Equinor’s geological data sources: Exploration and Production Data Store (EPDS, Equinor’s corporate data store for exploration and production data) and the NPD FactPages (a publicly available dataset published by the Norwegian authorities). Among existing challenges, the authors focussed on the limited flexibility of the data access points. OBDA is also a way of virtualisation of legacy data or data that must remain in its original form into highly scalable RDF schema.

4.2.5 The Siemens-Oxford Model Manager (SOMM) (energy sector/power generation)

A collaboration between Siemens and the University of Oxford was focussed on facilitating deployment of ontology-based industrial information models and explored the formalisation of information models using ontologies in two use cases in the manufacturing (ISA-88/95, three layers: product, process, execution) and power generation (IEC 81346 and ISO/TS 16952-10) domains. Their analysis revealed the need for integrity constraints for data validation. The outcome of their work was a tool named the Siemens-Oxford Model Manager (see Figure 10) aimed at engineers with little background on semantic technologies that partially implemented the OWL 2 RL profile extended with integrity constraints for data validation.

The work was driven by global initiatives to develop smart factories based on fully automated production processes and challenged by the difficulty of the seamless redevelopment and integration of industrial software components that had historically been developed independently of each other. The authors determined specific constraints that are needed in industrial use cases. Siemens’s R&D personnel is responsible for the development and maintenance of industrial ontologies used in the company. The goal of Siemens-Oxford collaboration was to widen the scope of ontologies applications to other teams of engineers. The proposed tool is based on an OWL (Web Ontology Language is a Semantic Web language) editor WebProtégé (Stanford University, 2021) and coupled with the rule inference engine IRIS (2021).

Property (*)	Required?	Min.	Max.	Range (*)	Value
<input type="checkbox"/> hasState	<input checked="" type="checkbox"/>	1	max	<input type="radio"/> State	Enter individual <input type="button" value="x"/>
<input type="checkbox"/> hasId	<input checked="" type="checkbox"/>	1	max	<input type="checkbox"/> xsd:string	Enter literal value <input type="button" value="x"/>
<input type="checkbox"/> hasConfig	<input checked="" type="checkbox"/>	1	3	<input type="radio"/> SteamTurbineConfig	Enter individual <input type="button" value="x"/>
<input type="checkbox"/> hasProductLine	<input type="checkbox"/>	min	max	<input type="radio"/> ProductLine	Enter individual <input type="button" value="x"/>
Enter property name	<input type="checkbox"/>	min	max	Enter datatype or class	Individual or literal

Figure 10. The SOMM editor

4.2.6 Dutch and Swedish National Road Authorities/TopQuadrant (transport sector)

This is a pre-commercial development driven by an EU R&D project and the Dutch and Swedish National Road Authorities called TopBraid Common Data Environment (CDE) (Topquadrant, 2021) based on the CDE concept found in industry BIM specifications. The Virtual Construction (V-Con) solution needed to support the Linked Data/Semantic approach and open information exchange/sharing between different stakeholders using various software solutions, tools and standards during the various life-cycle stages.

4.2.7 Highways England/BJSS (transport sector)

BJSS has designed and developed an ontology for Highways England by mapping the data entities and relationships held within the company. It has also produced a Proof of Value visualisation tool (BJSS, 2020) to ease the exploration and use of the ontology (see Figure 11). The data held by Highways England includes the Strategic Road Network with over the 4500 miles of motorways and major A-roads. Additionally, BJSS has developed a cloud-based data architecture platform using Azure and Databricks services to facilitate the building of data pipelines by other suppliers to provide access to difference data sources and enable connectivity across datasets.

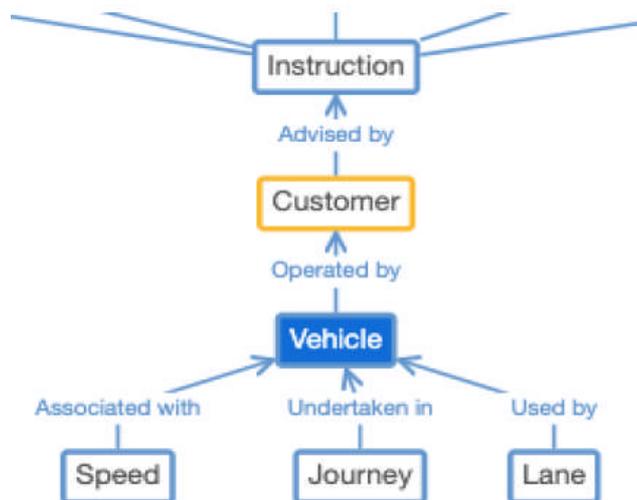


Figure 11. The terminology used by Highways England business units

4.3 Mid level ontologies

4.3.1 City scale ontology

iCity Ontology is an ontology for smart cities, to help and encourage those who shape policy to make decisions based on data. The iCity Ontology continues to be developed and recently presented a suite of ontologies for representing information used in transportation planning (Katsumi & Fox, 2019). Fox's earlier work includes a Global City Indicators Ontology, which integrates over 10 ontologies from across the semantic web, including geonames, measurement theory, statistics, time, provenance, validity and trust.

In the urban system, the following key concepts are defined: Person, Organization, Household, Building, Parking, Vehicle, Transportation Networks, Transit, Land Use and Travel. See Figure 12.

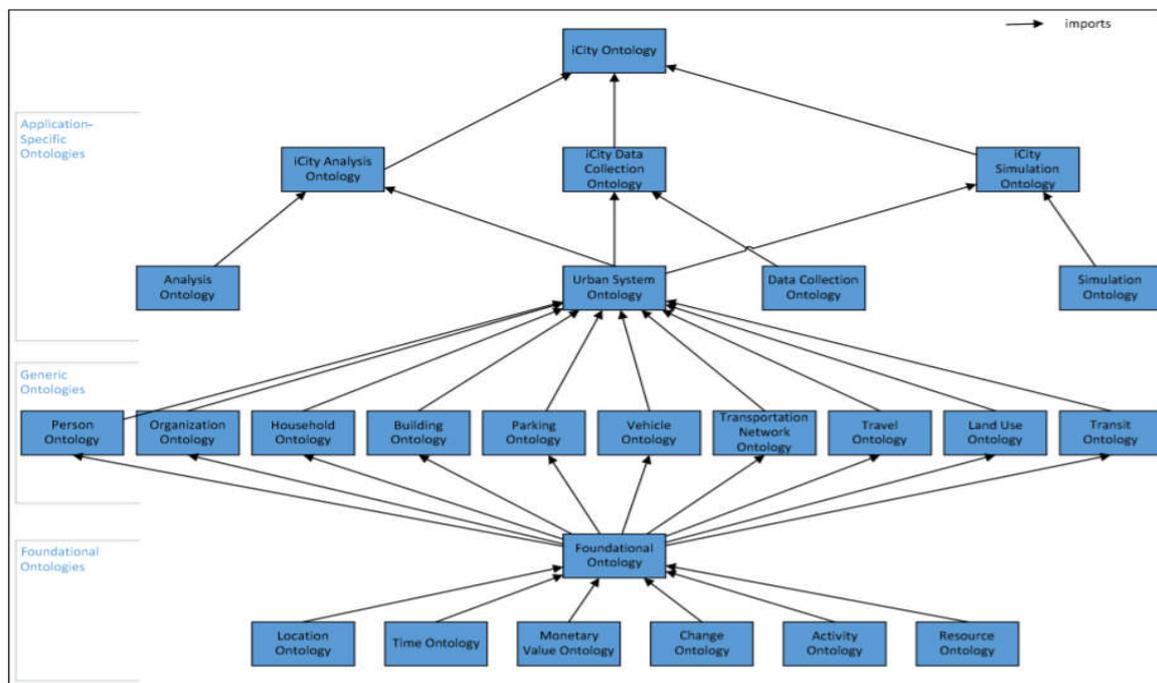


Figure 12. the iCity ontology (Katsumi & Fox, 2019)

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 yet aligned to an existing TLO: it is a mid level ontology without ontological commitment. However, it does leverage the key benefits of working with existing standards and taking a modular approach (4DSIG, 2020).

4.3.2 Digital Twin ontology

Whilst not in infrastructure, the work of Singh et al. (2020) proposes an ontology for the conceptual knowledge of the digital twin domain. The ontology represents three layers: physical, data and model, of the digital twin. See Figure 13.

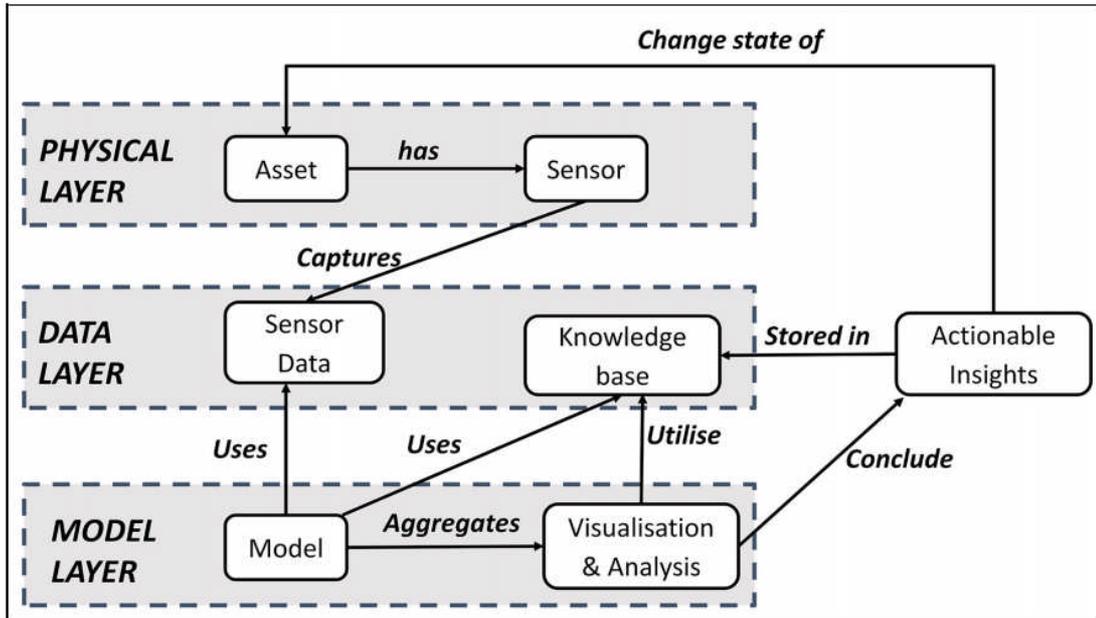


Figure 13. Digital Twin ontology (Singh et al., 2020)

Singh et al propose a simple process for the use of their ontology model to create and manage future databases for digital twins: Map, define, create, convert, and populate. The ontology classes are then connected to the functions of the digital twin. See Figure 14.

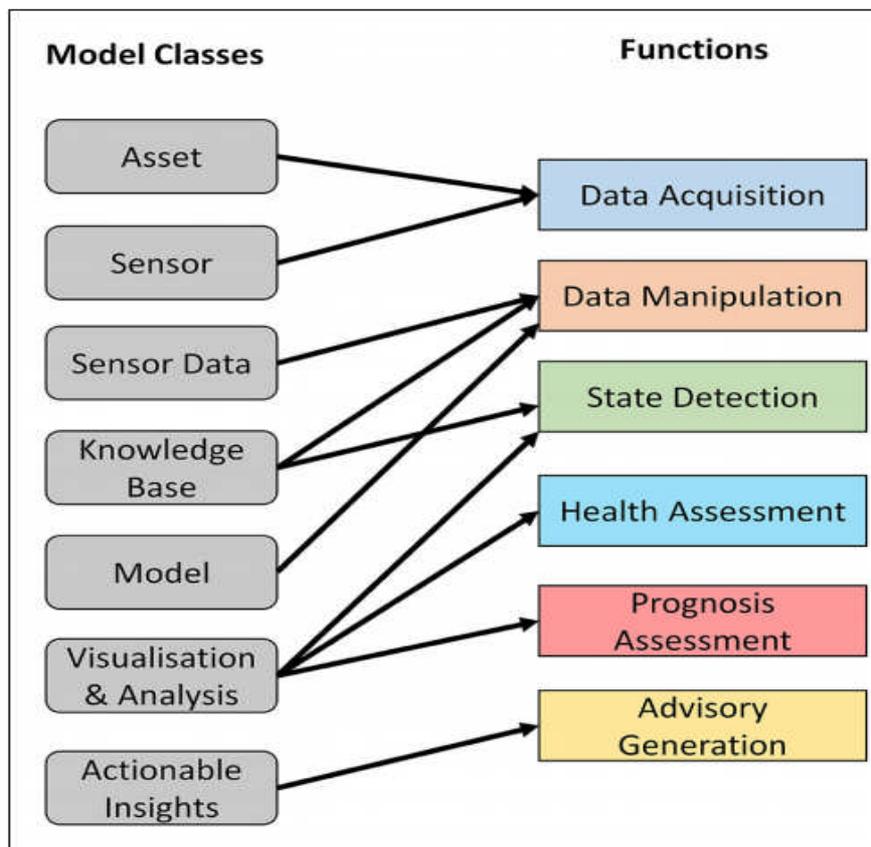


Figure 14. Ontology functional mapping (Singh et al., 2020)

The ontology might be considered as domain level, with application and sensor ontologies included. It does not include a top level which means manual methods for inferencing are needed. It is an important example because it connects the ontology classes to its aims (functions) so it facilitates functions, applications, use cases, etc.

4.3.3 Smart Cities Ontology for Digital Twins

In February 2021, Microsoft launched a Smart Cities Ontology, designed to align with their Azure Digital Twin platform. The Azure platform also introduced a new modeling language, Digital Twins Definition Language (DTDL), which can be used to describe twins in terms of the telemetry they emit, the properties they report or synchronize, the commands they respond to, and their relationship to other twins (Russom, 2021).

Microsoft's Smart City Ontology is DTDL-based, and utilises ETSI's Application Programming Interface Specification (ETSI CIM NGSI-LD), which defines an open framework for context information exchange (ETSI, 2021). There an information model that defines the meaning of the most needed terms, and a domain-specific extension to model any information. The core meta-model provides a basis for representing property graphs using RDF/RDFS/OWL, and is formed of Entities, their Relationships, and their Properties with values, encoded in JSON-LD. 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).

The contents of the ontology have so far been driven by an initial set of uses cases, which centre on the availability of IoT sensor networks. These cover the domains of environment, mobility, parking, and streetlight infrastructure., see Figure 15. Going forward, Microsoft plan to expand upon the initial set of use cases, and hope that modellers will be able to use both the Smart Cities Ontology and Azure platform to create digital twins.

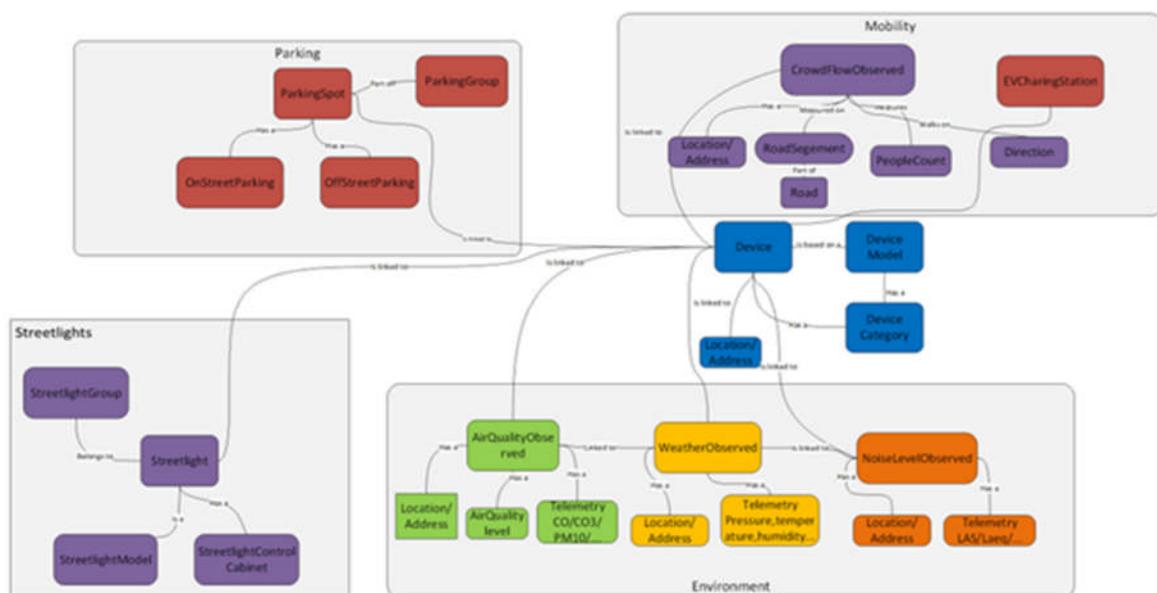


Figure 15. Azure Smart Cities Ontology for Digital Twins (Russom et al., 2021)

4.4 Barriers, risks, opportunities and strategies

An understanding of the barriers, risks, and opportunities for the adoption of digital twins is still emerging, and so strategies to overcome or take advantage of them are still vague. Motivated by perceived benefits of digital twins in the field of infrastructure, and some knowledge or experience of ontologies, several talks were provided before data was collected on 2nd Dec 2021 from over 100 attendees at a DAFNI ontologies and digital twins event. Three sectoral break-outs ran in parallel following three sectoral plenary talks. The key points from the sectoral discussions are presented in Table 2. Each sector raised points that demonstrated their distinctiveness however many cross-sectoral similarities appeared.

4.4.1 Cross-sectoral similarities

All sectors recognize the difficulties involved in tackling interoperability. These include gaps in knowledge, lack of knowledge to formulate business cases, and lack of incentive. Opportunities differ across sectors. Common risks include consequences of lack of capabilities and unknowns that might arise from adoption. Similar strategies include working together and learning from each other.

4.4.2 Energy sector

The energy (and water) sectors are especially concerned on privacy and security implications. Otherwise the technical challenges for data and interoperability in the energy sector and the great diversity of energy systems, diversity of users, and vast operational timescales are barriers. Opportunities for the use of near-real time data will help with decision support. Risks in the energy sector include adding overheads, creating new problems, and dealing with legacy technologies.

The shift toward distributed, disjointed, dynamic digital systems is evident in the energy and telecommunications sectors. When blending the old and new systems, challenges are intensified by mismatches in knowledge (Judge & Elahi, 2021).

Standards and frameworks in energy (and transport) sectors provide strategies for adoption but long established operational standards, processes and ways of thinking are barriers to innovation (Judge & Elahi, 2021). Governance and leadership as well as openness (for data and technology) are also strategies in the energy sector.

4.4.3 Transport sector

For the transport sector, barriers include cultural lock-in, and knowing where to start given the multi-scale and multiple 'truths' available. Opportunities raised for the transport sector highlight multiple methods that can be built upon, including a common data environment (CDE). In the construction, BIM (Building Information Modeling) includes a CDE "with explicit digital information and data exchange requirements, together with planned, structured information requirements at each project lifecycle stage" (Taylor, 2017, p56).

Risks include giving up the diversity of systems and the known ways of 'getting round' the systems. Other risks are the potential to overlook behaviours due to focusing too much on assets. Strategies embrace technology and automation.

Table 2: Sectoral barriers, opportunities, risks and strategies

Sector	Energy	Transport	Water
Aspect			
Barriers to ontology adoption	Gaps in knowledge; privacy and security implications; robust business model; hubris; technical challenges (data, interoperability); complexity (diversity, plurality, timescales, ...)	Too difficult: multi-scale; level of detail, finding the truth; lack of incentive: cost vs benefit; business case; gap in knowledge & skills; ability to learn value of data; where to start? Culture/lock-in; Trust and security; data-sharing	Lack of understanding; trade-offs; culture and governance; bureaucracy and silos; short-termism; lack of incentives; privacy, security and trust; not part of performance criteria
Opportunities	To integrate and share digital assets; measuring things not yet measured; use near-real time data; leveling up	Insight; Data quality; built on methods: MBSE (Model Based System Engineering) BIM, block chain/Ocean, ADEPT; B2B Common Data Environment	Hold data relevant to desired outcomes; balance stakeholders (including environmental); regulator must demand it; flexible and responsive to customers
Risks	Adding an overhead; data access; hidden assumptions; in-built bias; emergent problems; computer reasoning; social acceptability, legacy technologies	Systemic risks: feedbacks, too much focus on physical not behavioural; legacy; capabilities vary across operators; new standards may prevent 'hacks'	Financial and reputational; data verification; breaches; lack of skills/know-how; haven't learnt from BIM level 2
Strategies	Standards, frameworks, methods of unique identification, governance, leadership, data and technology openness and sharing	Technologies; collaboration; automation; learning from others; standards and moderators	Learn from case studies; trusted methods (e.g. data trusts); exploit sectoral strengths (low competition); commercialise digital twins; shared vision

4.4.4 Water sector

In the water sector, barriers include existing ways of working which are focused on the short-term, and targeting only those activities linked to performance criteria. Doing anything else would mean trade-offs. Opportunities to be flexible and responsive to customers would be created, but likely any action would need mandating by the regulator. Indeed risk management and ministerial invention are still the dominant barriers to radical innovation (Wagner & Fain, 2018). Risks include concerns that the sector has not yet learnt from BIM level 2. Strategies are perhaps unique in water given their monopoly. The opportunity to commercialise solutions is highlighted.

4.5 Digital Twins: capabilities and governance

The significance of ontologies is heightened in the context of connected digital datasets and solutions, especially digital twins and collecting of data from the real world. Opportunities for use of sensor data, integration with existing data, and with data across organisations, to solve use cases highlights some of challenges relating to capabilities and governance. Much of this will need to be automated and governed in new ways.

Data was collected via a Zoom Poll following talks on digital twins and ontologies in the infrastructure domain. The event was held on 02.12.2020. Over 100 attended and 66 useable responses were received.

4.5.1 Capabilities

The first question presented a variety of capabilities to make progress on a digital twin ecosystem. We were especially interested in the significance of ontologies, and put this capability 'last' to avoid selection bias (toward the first item) however the respondents attending were interested to learn about digital twin and ontologies and so bias was inevitable. We permitted multiple choice and results show that more than 4 out of 5 people marked "data conceptualisation and custodianship: ontologies, data trusts, etc." as important; more than any other capability. 30% of respondents selected all the capabilities, 10% selected three of them (always ontologies and behaviours – the first and last), 45% selected two capabilities, and 15% selected only one (half of these selected ontologies). See Figure 16.

Capabilities in data conceptualisation, custodianship, ontologies, data trusts, etc. are significantly important.

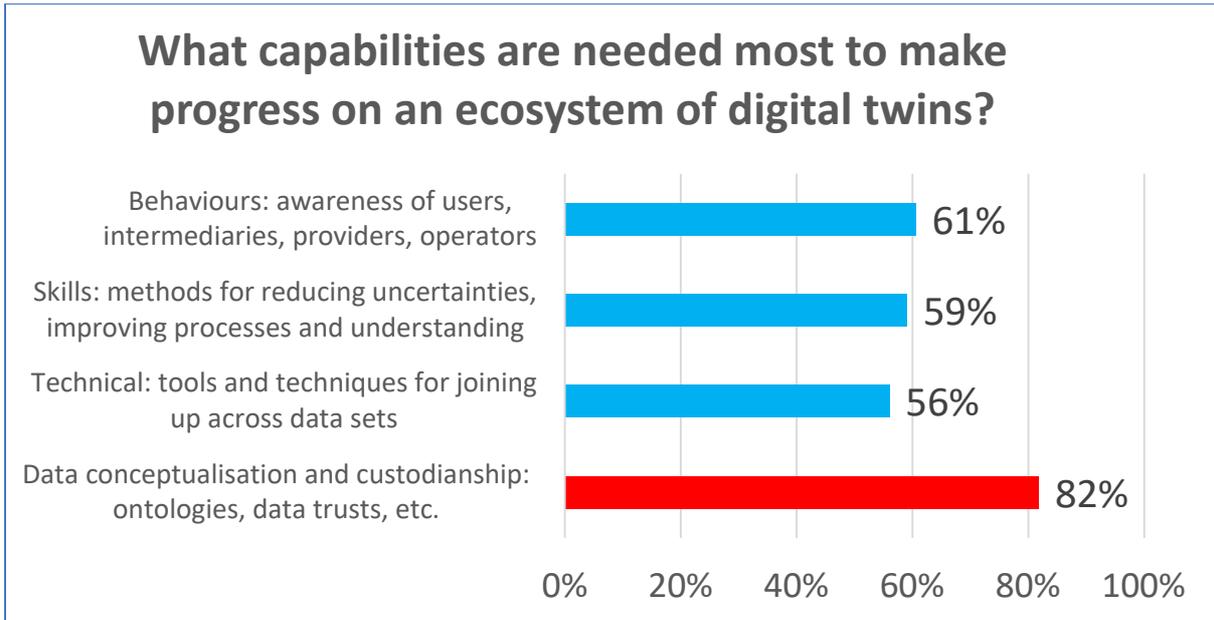


Figure 16. Capabilities to make progress on an ecosystem of digital twins

4.5.2 Approaches

The second question asked respondents to consider options for trialling possible solutions. For this question we were interested in where the emphasis ought to be: within domain/sector; across scale or across domain/sector.

All three approaches were rated highly, but the most popular was the Multi-scale approach. See Figure 17. This question was also multiple choice. 15% of respondents selected all the approaches, 35% selected two of approaches, and 50% selected only one.

Approaches to trialling should focus on multi-scale, and then on cross-sector use cases.

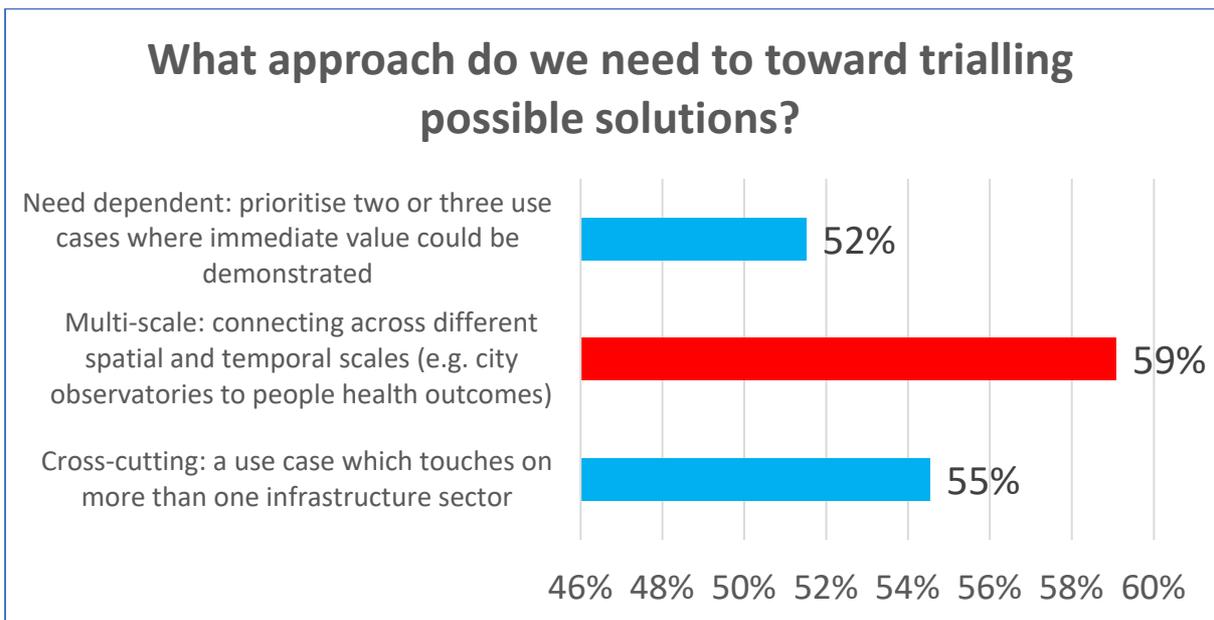


Figure 17. Approaches toward trialling possible solutions

4.5.3 Governance

The last question concerned options for governance of an ecosystem of digital twins. We suggested centralised, regionally organised, or organisation centric. The question was single choice and so forced respondents to identify their top priority. A regionally distributed governance regime was preferred. Just over one quarter wanted an organisation focus, the least popular vote. See Figure 18.

Governance needs to be distributed, but not at organisation level. It speaks to the Government's "Levelling Up" agenda and also demonstrates willingness to collaborate.

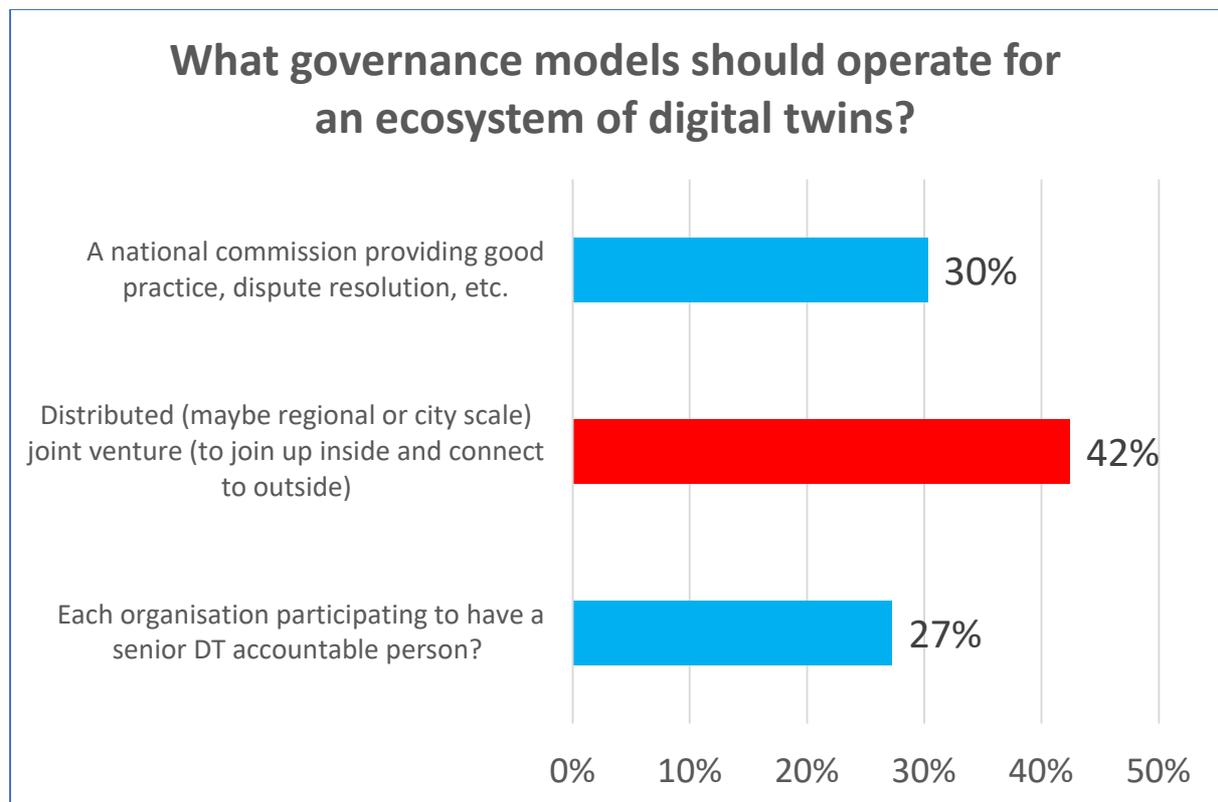


Figure 18: Governance models for digital twin ecosystem

5 Practitioners Panel Discussion

A number of important issues touching to top level ontologies are explored in this section. In order to achieve this, practitioners were interviewed, verification and feedback was sought after presentations, and the text iterated for comments. For each main issue identified, a take-away position is encapsulated in a greyed box. Interview and meeting dates are provided as citations and details of contributors are acknowledged in Section 7. Full interview transcripts are available on request.

For agreed and true knowledge of infrastructure systems with the purpose of exploiting real-world, right-time data for improved and responsive decision-making, a system of semantically correct mechanisms is needed for sharing of knowledge. The ability to share common traits of not only physical/built, but also environmental and social systems, requires that the data that is shared is characterised in an agreed semantically and explicitly defined way.

At the core of this is to find ways of aligning pluralistic claims to knowledge, i.e., to match the worldviews of one system, organisation, data provider, to any number of others, in ways that ensure the integrity of meaning behind any shared data. Assuming there is a desire to have consistency across infrastructure services, since they function effectively a system of systems for public good, then a hub and spoke model may work most effectively for conceptual definition. Inevitably the key discussion topics concern the need for a top level infrastructure ontology, which would enable the characterisation of knowledge, create interoperability, transparency, and have utility through agreed approximations of the real-world. In a top level ontology, different models or views of the world can be represented as different theories. Depending on the preferences of the user, specific theories can be used whilst still preserving the necessary assumptions of the top level ontology. Without a top level ontology there is an impoverished, and partisan conceptualisation: overcoming concerns on the feasibility of a top level ontology is a key pathway toward federated knowledge.

It is important to be clear about the scope of ontological framing. Ontologies are both computationally and technologically independent: they are agnostic of the computational means of implementation and of the actual technologies used. A top level ontology describes those components of infrastructure that requires common ways of definition, e.g. distances, times. The domain, application and other levels of ontology provide information of specific parts of infrastructure. For example, if water utilities want to share information on contaminated water with the government (and consumers), they will need a domain ontology that describes the qualities of potable water. The top level ontology may need the notion of quality but the domain level specifies the standard.

This section discusses builds on the state of the art review of the literature and on the industrial practices which have been discovered. It consolidates the insights and validates these with contributors and experts, in order to arrive at agreed knowledge which are presented in grey boxes in the sub-sections.

5.1 Interoperability

5.1.1 Plurality of top level ontologies

Different infrastructure systems have developed their own top level ontologies with wildly varying degrees of adoption. Integration of existing top level ontologies is unlikely to be

successful, due to mixed ontological commitments (see below for more on this) across sectors, e.g. possibilism, materialism (Partridge et al., 2020, p44). Finding a superior top level ontology will also be difficult because priorities and perspectives vary and will not be easily reconcilable meaning that a single authority will be needed for decision making.

Theoretically where multiple top level ontologies exist and are explicit (or self describing) then potentially multiple top level ontologies could exist although in practice they would merge into a single top level resource (4DSIG, 2020). Inconsistencies and incompatibilities would have to be resolved or removed (by a responsible authority) and a new top level ontology would be created for the digital twin ecosystem, for example a UK built environment digital twin ecosystem.

It would then be possible to integrate with other digital twin ecosystem which each have well-defined information management frameworks with foundation data models and explicit top level ontologies. For example, the UK built environment digital twin ecosystem could integrate with a UK natural environment digital twin ecosystem, or a French built environment digital twin ecosystem. An system of digital twin ecosystems is possible to allow exchange: however each new digital twin ecosystem dilutes benefits and replicates the costs, so a small number is desirable (West, 13.01.2021).

It will be impossible to mandate the use of a single top level ontology for everything in infrastructure, let alone everything in the built or engineered environment. This is even if one can be found that is superior.

Therefore a new top level ontology will need to be created, which is well defined and to which others can map.

5.1.2 Standardisation

If we want to communicate information we need to find a way to overcome ontological differences. One way is standardisation. An example comes from the Sensor Web Enablement (SWE) work at Open Geospatial Consortium (USA) (OGC, 2021) who state *“Standardization is the key requirement for communicating information about sensors and sensor data and for comparing and combining information from different sensors.”*

Standardisation works for mature domains where new datasets might adopt well-defined standards. It also works in industries where standardisation is mandated, for example, for compliance or quality assurance. However, for legacy data, the cost to standardise may be unjustifiable. Barriers may also exist, such as relinquishing control of in-house invented approaches.

Where standards (de-facto or explicit) are widespread for a particular purpose, they should inform ontologies: users may determine to opt out.

5.1.3 Ontological commitment

Each foundational ontology represents different ontological commitments. This means it aligns itself to a set of beliefs or paradigm or coherent set of theories.

A dataset commits to a set of things whose existence is acknowledged by a particular theory of system of thought (Partridge et al., 2020). If the top level ontology for a dataset is declared,

it is made explicit; if not, then it is implicit. Regardless the top level ontology is always applicable. “*Data structures and procedures implicitly or explicitly make commitments to a domain ontology*” (Chandrasekaran, Josephson, & Benjamins, 1999, p23). Furthermore, the strength of ontological commitment, the formal levels and universal levels of the ontology distinguish top level ontologies (Partridge et al., 2020).

For the Centre for Digital Built Britain, the distinctions identify the requirements and inform the ontological choices for a Foundation Data Model (FDM) for a federated National Digital Twin (CDBB, 2020). A consistent philosophical view is needed for the FDM. Mapping between things that are consistent is always relatively easy. Using something pragmatic, often mid level, is problematic. The FDM will have rigorous analysis methods, mapping into the reference data library consistently from source (West, 13/01/2021).

Organisations may take stock and decide to adopt the Foundation Data Model, Common Reference Data, etc. I.e. make applications native, and that would simplify their interfaces. For any new activity or projects, going native is better because it uses the most rigorous understanding of what is going on. Legacy is best left alone until it needs fixing, so you map from the legacy that needs integrating only (West, 13.01.2021).

Explicit ontological commitment is required for a dataset to determine the changes needed to the data to integrate with other data using a common top level ontology (TLO). The robustness of the proposed TLO of the FDM for the built environment is likely to make explicit ontological commitment desirable.

5.1.4 Matching at intensional or extensional levels

If we are to exchange data between digital solutions, the most costly route is peer to peer integration. This requires bespoke work and demands the reconciliation of different referencing and descriptive choices. The work is eased if two digital solutions have explicit ontologies which match or can be matched. Matching achieves semantic interoperability, essential for integration.

“*Ontology matching is a research area aiming at finding ways to make different ontologies interoperable*” (Schmidt, 2020). Existing frameworks and methods exist for ontology mapping and integration: in the Geographic domain Kavouras (2005, p189) identifies four types of integration: 1. Alignment 2. Partial compatibility 3. Unification 4. True integration. All four types of ontology integration deal with the conceptual (or intensional) level of the ontology and not the instances (or extensional) level. If we have only the extensional level (of instances) then there are questions regarding the suitability of the assumption that extensional information can be used as an inference mechanism for the taxonomic structure of the intensional level (Tomai & Prastacos, 2006). Both intensional and extensional may be present. Extensional approaches may be useful to test the TLO against, even though it is not sufficient on its own to set out the ontological commitments required to generate a robust TLO.

Most work on generating mappings between ontologies has focused on the intensional level of ontologies (i.e. on the level of concepts in the ontology) rather than the extensional level which is filled with instances. When mapping at the intensional level there is focus on what is needed to know in order to determine the reference of an expression (e.g. what is X, Or what it means to be X? Or can you define X?). Whereas extensional mapping attempts to identify the class of objects that an expression refers to (Tomai & Prastacos, 2006) which is more difficult than being provided with an intensional mapping, however extensional mapping may be an extremely useful and pragmatic way of developing a domain ontology and testing a TLO's fitness for purpose.

Extensional mapping may be available to 4D ontologies but at a representational level; extensional matching is a greater issue than 4D to 3D matching (4DSIG, 2020).

Matching ontologies via intensional matching is recommended. Data sets that require to join to the infrastructure research ontology will require the explicit creation of an ontology to enable intensional matching.

5.1.5 3D vs 4D world views

An important distinction (in the formal level stratification of top level ontologies) is between *endurant* and *perdurant* (colloquially referred to as 3D and 4D respectively). They have different perceptions of *occurrents* (things that happen) and *continuants* (things that exist) which highlight the consideration of the persistence of things over time (Hales & Johnson, 2003). 3D ontologies view individual objects as three-dimensional, with only spatial parts, and wholly exist at each moment of their existence. 4D ontologies see individual objects as four-dimensional, with spatial and temporal parts, and exist immutably in space–time.

In the survey conducted for this report five sixths of respondents indicated they had *endurant* (3D) rather than *perdurant* (4D) ontologies. The choice between 3D and 4D foundational ontology can lead to significant differences in the interpretation and comprehension of the conceptual models produced (Verdonck, Gailly, & de Cesare, 2020). Specifically 3D ontologies are understood more easily.

Nevertheless, there are some fairly standard mappings from *endurant* to *perdurant*, so this does not pose a major challenge to integration (4DSIG, 2020). “For example, assuming that the UK is using a *perdurant* ontology but, on a global scale, another nation may have used an *endurant* ontology for their national digital twin for the construction environment. There would need to be an interface between that whole digital twin ecosystem and the UK’s *perdurant* ecosystem and it would have a mapping.

The matching of 3D/*endurant* ontologies to a proposed 4D/*perdurant* ontology is not a barrier to inclusion.

5.1.6 Ontology consistency and quality

The need for interoperability points to a need to reduce inconsistency (e.g. between domains, systems, etc). This is where the use of a suitable TLO is important. TLOs 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).

In practice, choosing an explicit TLO is niche (i.e. not the norm). With multiple TLOs the cost of interoperability soars or is effectively impossible when matching cannot be achieved (see paragraph 5.1.4). However, the engineering analysis community does have data models which are implicitly 4D and West (2011) demonstrates how data models interface with ontologies (4DSIG, 2020). Engineering data models indicate a way forward.

Building on 4D industry data models could indicate how to create a top level ontology.

Another generally bigger problem is that it is unusual to see good ontologies. Data models often do not quite meet the requirements of the users, so users find creative ways to use a data model to meet their requirements by putting stuff in places that it doesn't quite belong. And so you need to recover that and understand what that requirement is as well (West, 13.01.2021). The consequence is that (existing) ontologies will have to be ignored, data will be examined and mapped to the integration ontology (West, 13.01.2021). Self-certification may indicate quality, as may 'fitness for purpose' (an engineering decision): it is about what works (4DSIG, 2020). But also hints that any TLO developed will need to have evolutionary mechanisms built in.

Applications are usually quite good in their core area of concern, but the further towards the edges you get, the flakier it gets. Those bits near the edges overlap with other areas. And there will be inconsistencies, because it's just not been done properly, and so that can easily be where a large source of problems are (West, 13.01.2021).

Inconsistency will be reduced through explicit creation of ontologies. Owners of data quality for any data to be shared across the infrastructure system will need engaging.

5.1.7 Automated reasoning

From an infrastructure research perspective, if we had ontologies then we could interrogate data across "*heterogeneous and disparate data- and knowledge bases*" (Hoehndorf et al., 2011, p1). Just as in bio-medical research, infrastructure ontologies, where they exist, "*do not sufficiently formalize the semantics of their relations and are therefore limited with respect to automated reasoning for large scale data integration and knowledge discovery*" (ibid). With ontologies the infrastructure systems community would advance data integration, automated inference and knowledge discovery.

Translating an ontology for the purposes of integration may be considered as the merging of the ontologies by taking the union of the terms and axioms defining them (Dou, McDermott, & Qi, 2005). Bridging axioms create bridges between terms in two related ontologies into a new ontology and for further merging. Although "ontology merging requires ontology experts' intervention and maintenance, automated reasoning by an inference engine can be conducted in the merged ontology in either a demand-driven mode (backward-chaining) or a data-driven (forward chaining) mode." (Dou et al., 2005, p2)

An illustration of the powerful automatic reasoning capabilities of an ontology-based framework (reflecting various aspects of the semantic relationships among the components in) can be found in the (intelligent) water distribution network (Lin et al., 2012). A complete test case scenario is provided to demonstrate the efficacy of the ontology-based service to assist automatic decision-making in cyberinfrastructure when failure occurs in physical infrastructure. Although a word of caution: ontologies built for reasoning generally speaking, require having to make so many compromises to enable the reasoners to work that your ontology starts to look less and less like the real world (4DSIG, 2020). Nevertheless, automated reasoning would be capable of managing semantic integration of big data through an inference engine (backward and forward chaining) even from extensional information rather than rely on human handling/'hackery' (Hetherington, 15.02.2021)

If automated reasoning capability is not a key driver of TLO adoption then the constraints on TLO creation are suddenly loosened favouring human understandability and navigability over first order logic robustness (McGee, 08.03.2021).

Automated reasoning should be investigated for its potential to integrate ontologies particularly across domains.

5.1.8 Ontological implementation barriers

Many ontologies exist in industry and academia: conceptualisations of the components of a cyber-physical system, application, domain, or mid level (such as a city) are widely available. They provide useful visualisations for shared understanding but are often ungrounded in top level ontologies. Mid level ontologies (which claim to avoid top level ontologies) create ambiguity as to whether objects are 3D or 4D. This opens up the possibility for different interpretations and introduces inconsistency (West, 13.01.2021). The absence of a top level ontology is just one of the reasons ontologies are not used for sharing data. Many other barriers were identified in paragraph 4.4.

The use of mid level ontology should be avoided..

A further barrier relating to sharing concerns the ability to uniquely identify resources. As data sharing will be over the semantic web, there is a need to provide Uniform Resource Identifiers (URIs). URIs may identify both physical and information resources (e.g. a URL is a typical URI for a web page). In order to reconcile an organisation's URI with another's (for the same thing) various methods may be used. These methods include linked data (as well as or to substitute for gaps in) or the analysis of extensionals (instances of data not described in an ontology). Linked data involves establishing equivalence (such as matching on date of birth) (Glaser, personal communication).

A method to reconcile the same things in different organisations is needed where ontological means are not available.

Investment cycles and capabilities are also barriers to information management progress. Although capital projects spearhead the use of ontologies, technologies and devices (OTDs), once capital expenditure is over such OTDs are not adopted into operational use. This is not only because whole life costs are not established at outset, it is also because organisations do not have the maturity to adopt novel OTDs. Knowledge and potential benefits are lost once project staff leave (Wray, 11.12.2020).

Organisations must invest in the whole life cost of innovation in information management, and develop their maturity to adopt and operate ontologies, technologies and devices.

Digital twins will create major benefits in operational settings, so investment into change must consider the cost of ownership, not just the cost of build. Cost of ownership must address building capability and skills in industry.

Organisations must reconsider the full cost of ownership to reap the benefits of digital twins.

5.2 Knowledge and evolution

5.2.1 Knowledge representation

Ontologies are at the most mature scale of knowledge representation: able to describe the most greatest degrees of internal complexity and expressive power of knowledge. See Figure 19. In particular axiom based ontologies that can deal with extensional constraints and description logics were most advanced and have developed from simple representations and specifications of knowledge toward semantic and structural understanding (LIU & ÖZSU, 2009).

Our ability to represent knowledge continues to mature. Developments such 'knowledge driven artificial intelligence (AI)' are including knowledge representation albeit only at level of incorporating triples into embeddings (so little/no ontological grounding) and the field would benefit from an ontological framework (4DSIG, 2020).

Axiom based ontologies provide the most mature means of knowledge representation as at 2009. Knowledge representation benefits from ontology engineering.

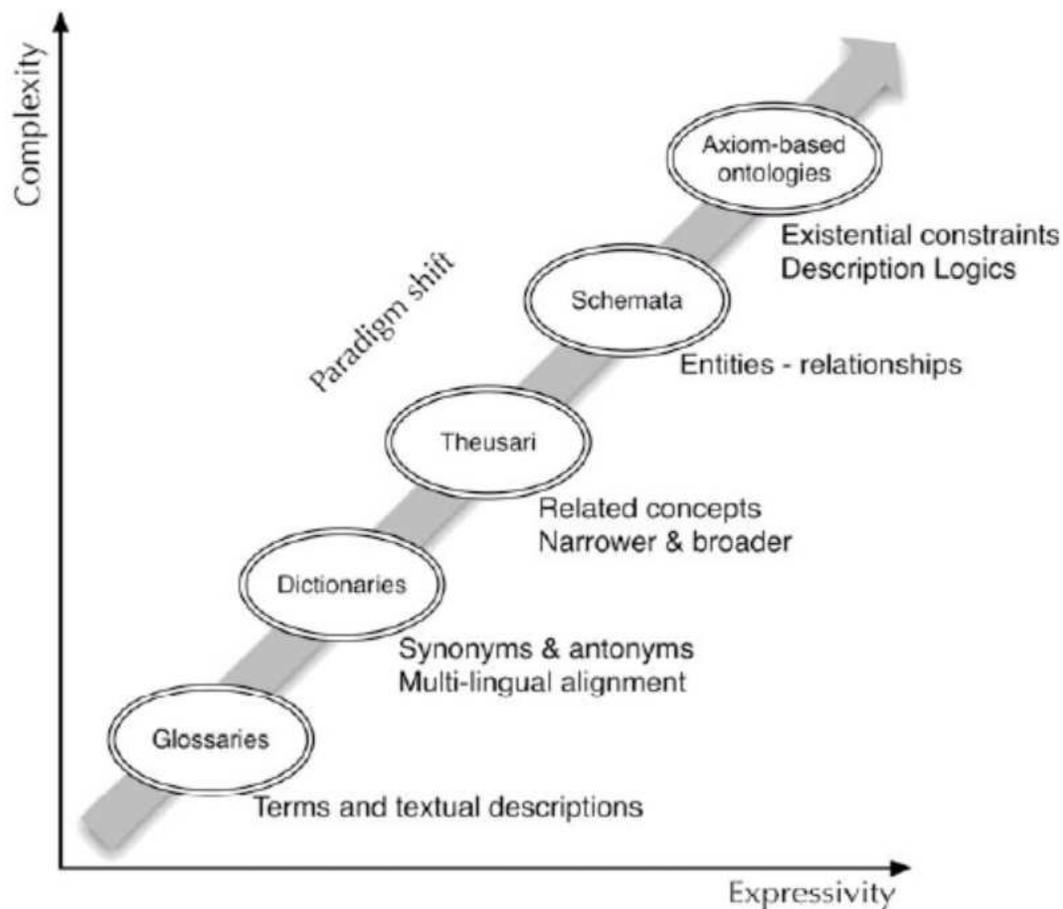


Figure 19. Knowledge representation paradigms (LIU & ÖZSU, 2009)

5.2.2 Coordinated evolution

It is safe to say that ontology development continues to mature. Some have proposed methods such as ontological realism for the coordinated evolution of scientific ontologies (Smith & Ceusters, 2010). Others have proposed methods for adaptation and change to ontologies and to manage meta-ontologies, see for example (Maedche, Motik, Stojanovic, Studer, & Volz, 2002; Stojanovic, Maedche, Motik, & Nenad, 2002).

Even Artificial Intelligence needs ontologies “*Any software that is useful cannot be written without commitment to a model of the relevant world*” (Chandrasekaran, Josephson, & Benjamins, 1999, p23). According to Žáček (2017) ontology engineering belongs to the realm of knowledge engineering which belongs to Artificial Intelligence. See Figure 20.

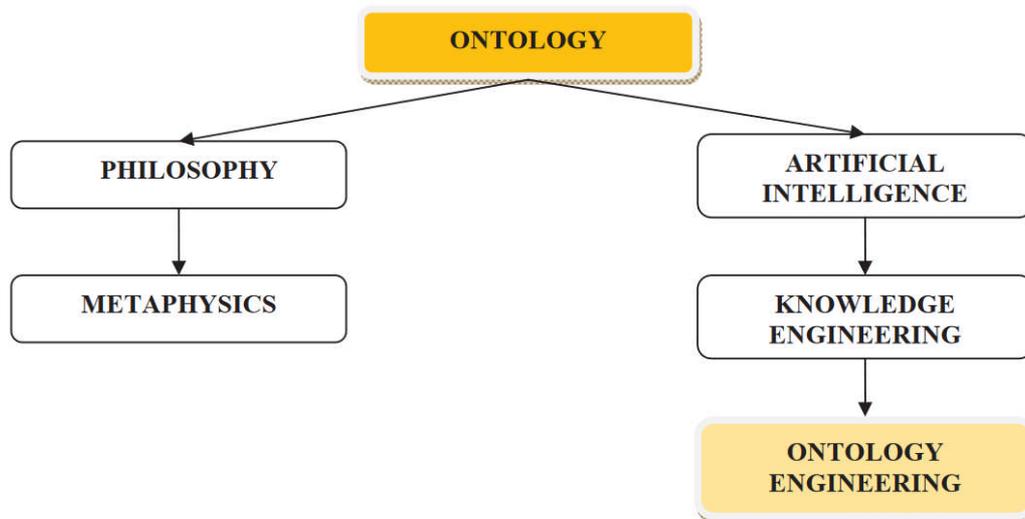


Figure 20. Ontology development (Žáček, 2017)

Mechanisms to evolve, replace, retire and introduce novelty into ontologies are needed and must be baked into standards.

5.2.3 Epistemology

Ontology is often contrasted with epistemology. Epistemology deals with the nature and sources of our knowledge (Guarino & Giaretta, 1995). Ontological development is without doubt concerned with ‘how we know what we know’. Some definitions of ontologies are couched in terms of epistemology and teleology (explanation of how goals are achieved): “*The key role of ontologies is to specify a data modeling representation at a level of abstraction above specific database designs (logical or physical) so that data can be exported ... and unified*” (LIU & ÖZSU, 2009) building on Gruber’s (1993) work.

Some have defined ontologies in respect of how we know them, i.e. from an epistemological perspective. Studer, Benjamins, & Fensel (1998, p184) state “*An ontology is a formal, explicit specification of a shared conceptualization*”. If we understand a conceptualization as a coherent sub-set of information, then it is useful to explicitly specify the ontology so that its boundaries and constraints are defined. If the ontology is formalized, it may also be machine

readable, enabling artificial intelligence mechanisms to infer information about the data sets to which the ontology relates.

Epistemology addresses questions on the truth of knowledge and related to that the methods of argumentation on truth value: see for example (Centre-for-Argument-Technology, 2014; Walton, Reed, & Macagno, 2008). Some ontologies also consider provenance: see for example Moreau & Missier (2013). Plato/Aquinas' transcendental values such as existence, truth, goodness/value need to be brought into the mix.

How we develop ontologies, what sources of knowledge we choose, how we interpret truth, must be in scope of ontology engineering.

5.2.4 First order and higher order logics

Arguably there is a layer above the top level ontology where you determine whether you are going with first order logic or higher order logic. First order logic deals with predicates (or objectives) and uses quantified variables to create expressions of logic or axioms. A theory may be expressed in first order logic. Only one semantics is studied.

Second and higher order logics allow predicates and higher level relations to be quantified introducing the capability to have several possible semantics or full semantics (D. Miller, 1991). This makes it more expressive, but with higher order logics, there is no effective deduction system.

5.3 Use cases

5.3.1 Background

Use cases are used to capture functional requirements and to understand the interfaces between people and use cases in a system of interest (e.g. Moreira, Araújo, & Brito, 2002). Use cases signal the need for new datasets and processes, providing leadership for change and improvement.

It is worth mentioning here that the NDT provides the plumbing that gets the data from source to destination and ensures the provision of a consistent dataset. So use cases and apps are not part of NDT but likely to be a whole market of applications that people can use (West, 13.01.2021). The creation of virtual integrated data stacks may be found useful for purposes other than primary use cases. For example there is great potential for DAFNI to use consistent datasets provided by the NDT (Enzer 13.01.2021).

5.3.2 Types of use case

There are two generic kinds of use case: the first is like the Grenfell Tower case, where there is the same kind of data about different objects which is held by lots of different organisations. It needs to be brought together to do something useful with the whole. The other is where there are overlapping data sets (about the same things/object). Examples include: emergency response, where you need to bring data together across different services and utilities, and

transport planning, where multi-modal transport where it is quite hard right now because the data sets are inconsistent (West 13.01.2021). It is possible to think of a third type which is a hybrid of both.

Two other alternative types of use case are possible. Both consider the systemic effects of data incompleteness or inaccuracy. The first is where one or more sub-systems are performing sub-optimally, which may lead to systemic failure. The second concerns the uncertainties of future interventions which require assumptions about data because they cannot be known in advance.

5.3.3 Use case: smart motorway control of autonomous vehicle speeds

Background

The use case developed for this IRO project would fit into the latter type of use case and it concerns road safety in the context of emerging autonomous vehicles using smart motorways.

In the transport sector, legacy ontologies are siloed and operationally focussed (e.g. using sensors, control rooms), with tight organisation governance with a central planning mind set. Newer ontologies are emerging and can be described as information technology driven, with decentralised IT, using Internet of Things concepts, and taking a 'data as a service' approach with many cultural differences from the 'old world'.

The consequence is evident by established automotive manufacturers who struggle to integrate with transport system data and to keep up with the speed of technology change. This compares with companies like Tesla who have a digital mindset, have designed for over the air upgrades, and transport authority and smart motorway schemes integration.

Smart motorways were introduced to increase road network capacity by use of variable speed limits to stop bunching and congestion which improves safety by reducing accidents, reduces journey times by smoothing traffic flow, and cuts particulates by gentler braking and accelerating behaviours. Smart motorways use the hard shoulder as well as gantries which appear at 500 metre or 1000 metre intervals with electronic displays of speed limits.

At present only 36% of drivers notice the gantry information because it is so similar and frequent. The result is a great diversity in responses to notified speed limits which creates variety in the dynamics of traffic flow.

Challenge

In 2022 when speed limiters will be introduced into vehicles, all vehicles integrated with smart motorways will autonomously conform to speed limits. This is just another example of the generative power of the internet (Zittrain, 2009). Enabling successful integration between transport, energy, telecoms and automotive systems requires conceptual clarity at the intersection of mobility and infrastructure systems: data streaming (e.g. Android Auto), vehicle platform (automotive OEMs), smart motorways, and most likely energy services (e.g. for electric vehicle mobility).

The reality is that different standards exist for different communities with no obvious integration route. Each community tends to have a very poor understanding of the other.

- PAS1883 is the third publication from a series being supported by the Centre for Connected and Automated Vehicles.

- CCAV is very much focused on the vehicle side, as opposed to infrastructure: a major flaw in the CCAV set up and objectives as, almost by definition,
- ISO14812 has almost the exact opposite problem. It is very much from a highways infrastructure perspective.

New standards and technologies (some mandatory and other optional) continue to emerge as autonomous vehicle and information systems continue to mature as do intelligent transport systems. See Figure 21.

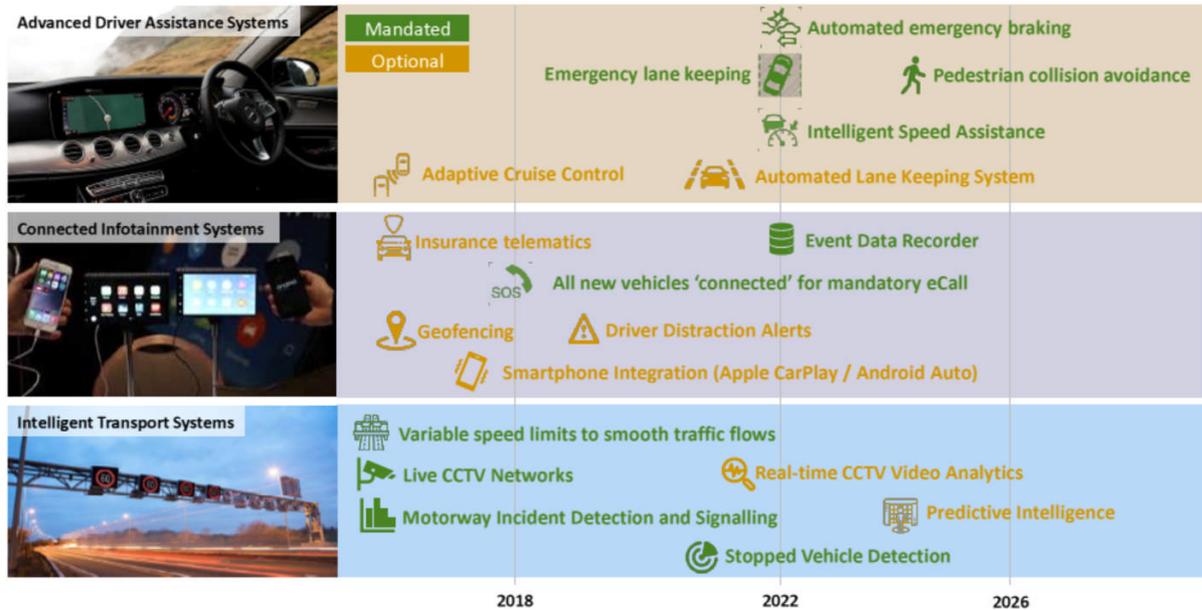


Figure 21. Roadmap showing mandatory and optional standards for autonomous vehicle and intelligent transport system © John Batterbee, Costain

Furthermore, the effects of desired outcomes of smart motorways (and on driver experience and environmental impact) are largely unknown, especially considering that the proportion of vehicles with speed limiters will be small but growing over the coming years. A means to automatically collect and analyse data on the effect of speed limiters is highly desirable, in order to improve decision making.

5.3.4 An ontological approach for smart motorway control of autonomous vehicle speeds

Domain ontologies

Bagschik et al (2017) review ontologies as knowledge-based systems in the field of automated vehicles. Their aim is to generate knowledge-based scenarios which provide a way to assess if automated vehicles can fulfil a safe driving task, and comply with the hazard analysis and risk assessment demanded by the ISO 26262 standard (ISOTC22/SC32, 2014).

Published as a W3C Draft, the Road Accident Ontology (Dardailler, 2012) describes traffic and road accidents involving people, vehicles, animals, and relevant information, such as location, cause, involved parties, and so on. The ontology has not been applied in practice (Katsumi & Fox, 2018).

An older effort on Ontology for Transportation Networks (OTN) was created based on Geographic Data Files (GDF) (Lorenz, 2005).

To reach full autonomy of road vehicles it is necessary to provide an accurate and comprehensible situation description for the environment and the vehicle itself. A consistent depiction is essential to facilitate data exchange and communication including with infrastructure, e.g. smart traffic lights, road signs or radio traffic service (vehicle to infrastructure) (Brunner, Kucera, & Waas, 2017).

Unmanned vehicles must collaborate across multiple domains/environments (air, surface, land, underwater and space). There must be semantic integration of the sensing data into a single common data architecture with agreed data structures and APIs, supported by a common semantic model to ensure shared semantics. The scope of such a model extends beyond the APIs, as their semantic integrity depends upon the systems behind them respecting (and so understanding) it. The constructional approach to ontology development supports the semantic unification requirements of multiple platform-domain systems, such as unmanned vehicles (Partridge et al., 2019).

A National Digital Twin solution

The information management framework to solve the use case would require a variety of components, described below.

The foundation data model (containing the top level ontology and the information management framework) will provide the data structure (West 13.01.2021). Reference data, which is mostly static and tends to be classes, is part of the data structure. Shared reference data would include item such as urban speed limits, roads, and other asset data. Master data varies more frequently, for example, as speed limits change at a particular location. A thin slice refers to the ontological classes that provide instances of master data. A thin slice is the smallest thing that does something useful and has a business case in its own right (West 13.01.2021).

The National Digital Twin of the built environment would recognise the authority of particular agents, for example, permitting only valid thin slices of variable speed limits, to ensure the security of the system. Specifically where this validation is conducted must be described in the information management framework.

Any type of autonomous vehicle within range of the information in the thin slide would receive the thin slice in a format interpretable by the information management system of the autonomous vehicle. This could be done only if the top level ontologies of both the National Digital Twin of the built environment and that of each autonomous vehicle manufacturer is known. 4D ontologies are very well equipped to deal with this particular use case: variable speeds will be notified to those in the smart motorway ecosystem based on space/time, i.e. a particular gantry at a particular time. Inherent capabilities of 4D ontologies to deal with space/time are well suited to time varying data (such as variable speed limits between motorway gantries). In 3D ontologies more complex handling is needed.

In addition, the thin slices can be amalgamated in virtual databases which can be analysed using machine learning to determine traffic dynamics, environmental conditions, risk of road accidents, etc. This data would inform the setting of future gantry speed limits, but also to inform future transport operational and planning decisions.

5.3.5 Use case reflection

The role of the national digital twin vs the local digital twin

This use case brought to us by Costain is an excellent demonstration of the need to integrate across sectors, and across scales. It also highlights that digital twins need to consider timing: autonomous vehicles need to be informed at the 'right-time' about the upcoming change in variable speed limit(s).

We assume that Highways England, or the smart motorway operator will (in order to determine to variable speed limits) collect data on the levels of congestion, accidents, air quality, etc. of the transport system. It will then intelligently extrapolate (using some model) likely congestion using also planned roadworks, pre-existing knowledge on times of congestion, etc. and then recommend to a decision maker (or increasingly in the future) it will actuate variable speed restrictions.

Thus the operator will have a digital twin in order to determine speed limits. A national digital twin will allow the variable speed limits to be communicated to car users, whose speeds and behaviours will be detected by the operator's digital twin. The operator will be able to quantify the efficacy of variable speed limits.

The need for appropriate information storage and transport protocols

The solution requires consideration of storage and data transport mechanisms. How quickly will autonomous vehicles be notified of variable speeds? Will data will stored locally (distributed) and where? (Or perhaps there an implicit assumption of reliable, centralised storage with high-speed transmission to those who must have the data? But will this scale and does it create too much risk of single point of failure?)

What data storage formats will facilitate rapid processing? What telecoms transport routes and protocols will be used? How will locations of moving autonomous vehicles be determined especially given interruptions to 3/4/5G networks?

Storage and transport mechanisms will likely vary across digital twins: they will need to be appropriate for the use case in hand.
--

6 Recommendations, assumptions and next steps

Recommendations for next steps are drawn out from literature review, data collection and the discussion. Whilst an attempt to prioritise recommendations was made at a separate project event (29/01/2021) there was very little to differentiate them, although a wide range of assumptions were identified. Recommendations are listed in paragraph 6.1, assumptions in 6.2 and proposed next steps in 6.3.

6.1 Recommendations

- R1. Practice 'dog-fooding' (using ontologies ourselves) in particular on a cross-sector, cross-scale use case with industry input
- R2. Conduct explicit mapping of infrastructure use cases to ontological levels, scale, etc.
- R3. Consider explicitly defining geospatial schemes (e.g. NUTS) and semantic sensor network architectures for infrastructure models using digital twins
- R4. Do not attempt to convert legacy models for inclusion in an infrastructure ontology (instead focus on datasets that are input to the legacy models)
- R5. Do not accept data for sharing unless the ontology of the dataset is provided (or a method to reconcile the same things in different organisations is available)
- R6. Address some of the gaps in knowledge by providing information on known ontologies and digital twins
- R7. Align to the CDBB Top Level Ontology and Information Management Framework
- R8. Use 4D industry data models to inform a top level ontology
- R9. Support organisations to build their maturity to adopt and operate conceptualisations, technologies and devices

6.2 Assumptions

It is important to note that assumptions are often implicit in contemporary narratives. To them explicit we brainstormed our assumptions. Specific assumptions were noted, and for the purposes of this report, are clustered by similarity. Assumptions are also brought forward from the discussion section.

6.2.1 Purpose of infrastructure ontologies and digital twins

It is assumed that:

A1.1 The **core principles** (security, trust, etc.) to be met by ecosystems in which ontologies are used **can be explicitly identified** (paragraph 3.2).

A1.2 Consensus exists that **a shared infrastructure top level ontology** for infrastructure systems would avoid costly peer to peer integration. Once a data set's ontology is made explicit, it can be used to match to the shared infrastructure top level ontology and re-used. Federation **is the lowest cost model** for sharing (paragraph 3.1).

A1.3 The role of use cases is fundamental to the improvement of infrastructure services and to achieve desired outcomes. **Use cases**, including the use of personas, UX – user journeys, etc., **will drive the development** for infrastructure ontologies.

A1.4 **Findings** of infrastructure models **would be improved** with data from infrastructure operators.

A1.5 **Models** of UK infrastructure **would be more accurate if they used the ontology for the infrastructure domain**.

A1.6 **Use cases for DAFNI platform are NOT the same as use case for infrastructure improvement**.

6.2.2 Data access and sharing

It is assumed that:

A2.1 Organisations, institutions, and companies working in infrastructure will be **willing to (or compelled to) share their data**. Regulation can enforce this but we have an eclectic mix of private and public enterprise, joint ventures, and many forms of business model. There are implications for sub-contracting and data protection, amongst others.

A2.2 There is **gap in data sharing knowledge** in infrastructure systems that can be filled by **learning and transfer from other industries**: we don't necessarily need to learn from scratch on data sharing. E.g. Open Banking has common data sharing (computer to computer), not just common data environment (Starks, Cheetham, & Patchay, 2020).

A2.3 The **same data will be shared as that used internally**. The assumption is that the effort to share is minimal because data does not need to be changed significantly.

A2.4 **Mechanisms exist for handling data privacy, licencing**, etc. such as data trusts.

A2.5 **Sharing will improve data quality** and that eventually there will be a high quality dataset that can be used to integrate most other things, some single source of truth. This is consistent with "Inconsistency will be reduced through explicit creation of ontologies (paragraph 5.1.6).

A2.6 **Explicit ontological commitment** is required to integrate datasets using a common top level ontology (TLO) (paragraph 5.1.3). This means intensional matching can be used (paragraph 5.1.4). Also that "automated reasoning should be investigated for its potential to integrate ontologies particularly across domains" (paragraph 5.1.7).

A2.7 "**Storage and transport mechanisms** will likely vary across digital twins: they **may need to be appropriate for the use case in hand**" (paragraph 5.3.4).

A2.8 **Owners of data quality** for any data to be shared across the infrastructure system **will need engaging** (paragraph 5.1.6).

6.2.3 Approaches and method for creating ontologies

It is assumed that:

A3.1 It is **better to create an ontology from data** (bottom up) rather than creating abstract models and trying to fit to data (top down). “The top level ontology will need to be well defined and to which others can map” (paragraph 5.1.1). This also means avoiding mid level ontologies (paragraph).

A3.2 The **robustness of the proposed TLO of the FDM** for the built environment **is likely to make it desirable** (paragraph 5.1.3).

A3.3 Selected/created **ontologies will match and align with how different 'scales'** of application and sharing **are used in reality**. This is a usability assumption.

A3.4 **Methods in infrastructure construction**, e.g. use of concrete and steel, **will not shift dramatically**, e.g. 3D printing of concrete is not reflected in current ontologies. So the top level ontology devised for the built environment is not expected to change much.

A3.5 Ontologies **can support the explicit identification of scales** (temporal, spatial, and geographical) but further understanding of the underlying processes is needed to aggregate, disaggregate, and connect datasets (extents of points in space and time).

A3.6 **All information objects within a given space** will be assigned to an ontology item, if they are in scope of use cases for infrastructure knowledge engineering.

A3.7 For **infrastructure modeling** we may need to add abstractions for scenarios / possible futures, abstractions for model inputs/outputs, data verification/quality, etc. to our ontologies.

6.2.4 Approaches and method for implementing ontologies

It is assumed that:

A4.1 **Ontologies can be developed into schemas** and taxonomies, e.g. schema.org.

A4.2 Vocabularies and taxonomy **standards can be found that permit alignment of broad and varied data themes** (infrastructure plus wider themes, e.g., environmental). EuroSciVoc is one such potential taxonomy.

A4.3 “Where **standards** (de-facto or explicit) are widespread for a particular purpose, they **should inform ontologies**: users may determine to opt out” (5.1.2) The role of taxonomy is to harmonize/align original sources (provenance).

A4.4 **JSON Linked Data has a role in serialisation**, but may not be so useful for top level expressibility. Assumes JSON handles predicate properties but has freedom to store different structures inside. Use CSGL for extensions

A4.5 FAIR principles with DCAT+ may be used for **data discovery** (but won't address everything).

A4.6 DAFNI-LD (Linked Data) is extendible. Not only we can add specialist ontologies but structurally meet new requirements e.g. NUTS level derived from GeoJSON.

A4.7 **Formal standards will be selected** - e.g., DCAT-2 complemented by resolutions from GeoDCAT-AP for geospatial entities (e.g. metres, scale, angular) and ISO 19115 for use with DAFNI-LD.

A4.8 Model developers will be motivated to **annotate their models with ontologies** / explicit metadata.

A4.9 DAFNI will be able to **differentiate between data that has ontological commitment and that which doesn't**.

6.2.5 Mechanisms for ontology evolution

It is assumed that:

A5.1 Ontologies and ontological paradigms will **continue to evolve** (paragraph 5.2.1).

A5.2 **Mechanisms to evolve**, replace, retire and introduce novelty into ontologies **are needed** (paragraph 5.2.2).

A5.3 How we develop ontologies, what sources of knowledge we choose, how we interpret truth, **(epistemology) must be in scope of ontology engineering** (paragraph 5.2.3).

6.3 Next steps

DAFNI is in a very strong position to lead on the development of infrastructure ontologies and digital twins, given the UKCRIC family of facilities and urban observatories (in addition to DAFNI's compute and platform capabilities). The agnostic role of the Science and Technology Facilities Council and co-investigators and teams with leading track records in national infrastructure modeling, data, analytics and visualisation, provide an ideal environment for taking forward ontologies and digital twin investigation.

In order to make a start on the very long journey of ontology adoption for data sharing and digital twin development, some first steps are proposed below.

6.3.1 Action 1: Create demonstrators

The best way to engage and demonstrate the potential for success is to create one or more demonstrators which targets the assessment of a particular use case but also opens the door to future development. As set out in the recommendations, the selected use case needs to be cross-sector and cross-scale. The use cases must select a top level ontology (allied to CDBB's TLO) and must deal with the selection of storage and transport mechanisms for shared data. Some options for use cases are suggested in Table 3, and include the current project's use case (autonomous vehicles and smart motorway integration).

Table 3: Use case options

Type	Examples	Use Case
(i) instances/utilisation of the same materials, methods, products, ...	Grenfell Tower cladding	I'm at Highways England and a road bridge has unexpected fissures seemingly from the materials used in construction. Would like to raise a product alert, get some empirical research done on the bridge, get the government to commission an assessment of the scale across the country, and get 'vaccine' in place to remedy.
(ii) event with potential to cripple multiple services	Lancaster Flooding	I'm at the Environment Agency (EA) with knowledge of a specific flood risk following long-term Met Office weather predictions. EA would like to share details of the area of potential impact with infrastructure operators and agencies, to allow assessment of risks to assets and service users. A collective assessment quantifying the risk to resilience of the infrastructure services in the area and the likely impact on people, commerce/industry, assets, the ecology and wider environment, enables the prioritization of interventions to maintain resilience.
(iii) suboptimal services creating risk of systemic failure	Hidden Hazards	I'm at the National Grid and introducing syngas into a part of the network. Supply changes have created a risk of some components failing and having rebound effects. Would like data to be collected across the distribution network and customers locations, to assess safety risks and the potential for the whole system to fail.
(iv) unknown future intervention effects	Variable highway speed limits for autonomous vehicles	I'm at Mercedes and introducing autonomous vehicles with speed limiters. I want our driverless cars to safely meet variable speed restrictions provided by Highways England at regular highway gantries. Would like to collect these speed limits as they occur and issue the right vehicles with instructions to change their speeds (safely) and provide information to Highways England on the vehicle dynamics triggered. The data can be used by Highways England to assess road safety for the speed limits at the following gantry.

6.3.2 Action 2: Fill knowledge gaps

Review and upload details of ontologies discovered in this research onto the DAFNI website and platform, creating (a) ontologies for infrastructure and (b) digital twins for infrastructure. Provide seminars, training and other forms of professional development, targeted at industry, to build awareness, confidence and the skills to develop business cases and solutions that provide value and infrastructure-wide benefits.

The DAFNI platform is particularly well placed to host ontology and digital twin sharing, as it already provides data, models and visualisations for infrastructure research.

6.3.3 Action 3: Build a network of collaborators

Academics, even modelers and those very familiar with modeling and data management methods, will not be sufficient to avoid pitfalls. A network, already started as part of this work, needs to be developed, to provide specialist skills (e.g. security, privacy, legislation), challenge, review, ideation, and generally help to manage uncertainties for the purpose of better infrastructure services.

6.3.4 Action 4: Co-create new governance and business models

Governance will likely need to be distributed at a meso-level above organisations. Exploration is needed on how exactly this will work. This will involve stakeholder engagement to co-create the governance and business models, develop a road map for infrastructure systems, and a long-term programme of work.

6.4 Gaps and limitations

Aside from the areas for action mentioned in section 6.3, there are gaps and limitations which could extend this work into other systems of interest or could delve deeper into other characteristics of infrastructure ontologies.

The system of interest for this paper was infrastructure, i.e. transport, energy, waste, waste and telecommunications systems. Others can take this work forward by including other systems into scope, such as buildings, manufacturing, rivers, etc. The work here also targeted information sharing and interoperability, and so has focused on top level ontologies. Although it did identify other levels of ontology in infrastructure systems, further investigation may discover other useful characteristics. Otherwise, ontologies are continuously being developed, as are the information management frameworks in which they reside, as their value is increasingly acknowledged. For infrastructure research, they are yet to be exploited but there is a growing body of evidence that urban observatories and asset sensing, are leading the way with near real-time information, would benefit from ontologies, and would revolutionise the value of models which have previously had to manage without real data. Researchers have the opportunity to consider how to subsume typically ex-ante models in a new era of digital twins.

7 Acknowledgements

We thank the DAFNI governance board for approving the champions initiative, and ultimately for EPSRC funding from grant EP/R012202/1. Furthermore we gratefully acknowledge the participants in the project who gave freely of their time. Without them we would not have created such a rich insight into ontologies for infrastructure research. We classify participants by their primary activity.

For interviews

Mark Fox, University of Toronto
Sergio de Cesare, University of Westminster
Matthew West OBE, CDBB
Hugh Glaser, Seme4
James Hetherington, CDBB
Simon Wray, ThinkSPI

For reviews and comments

Chris Partridge, Chief Ontologist at BORO Solutions Limited
Pierre Grenon, Principal Ontologist at BORO Solutions Limited
Andrew Mitchell, Principal Ontology Consultant, also BORO
Liam McGee, Chief Scientist and Founder at Kaiasam

For presentations

Phil Young, Digital Catapult
Simon Evans, ARUP
Volker Buscher, ARUP
David Wyatt, Energy Systems Catapult
Ian Gordan Highways England
John Batterbee, Costain
Ben Kidd, ARUP
Brian Matthews, DAFNI

For industry engagement

John Batterbee, Costain
Mark Enzer, OBE, Mott Macdonald
Ben Kidd, ARUP
Phil Young, Digital Catapult
Liam McGee, Kaiasm

For DAFNI support

Marion Samler, STFC
Rocio Garavito, STFC

For contribution to the ontology survey and participation at events:

Too many to name - thank you!

8 References

- 4DSIG. (2020). 4D SIG meeting 11.12.2020. *11 Dec 2020 Meeting*.
- 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>
- Adida, B., Birbeck, M., McCarron, S., & Herman, I. (2015). RDFa Core 1.1 - Third Edition: Syntax and processing rules for embedding RDF through attributes. Retrieved March 5, 2021, from <https://www.w3.org/TR/rdfa-core/>
- AGI. (2015). (Association for Geographic Information), UK Gemini. Retrieved March 4, 2021, from <https://www.agi.org.uk/agi-groups/standards-committee/uk-gemini>
- 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
- Andrews, H., & Wright, A. (eds). (2019). JSON Hyper-Schema: A Vocabulary for Hypermedia Annotation of JSON. Retrieved from <https://json-schema.org/draft/2019-09/json-schema-hypermedia.html>
- 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>
- Arp, R., Smith, B., & Spear, A. D. (2015). *Building Ontologies with Basic Formal Ontology*. MIT Press.
- Ashino, T. (2010). Materials Ontology : an Infrastructure for exchanging materials information and knowledge. *Data Science Journal*, *9*, 54–61.
- Atkinson, R., García-Castro, R., Lieberman, J., & Stadler, C. (2017). Semantic Sensor Network (SSN) Ontology. Retrieved March 2, 2021, from <https://www.w3.org/TR/vocab-ssn/>
- Bagschik, G., Menzel, T., & Maurer, M. (2017). Ontology based Scene Creation for the Development of Automated Vehicles. *ArXiv*, (lv), 1813–1820.
- Bajaj, G., Agarwal, R., Singh, P., Georgantas, N., & Issarny, V. (2017). A study of existing Ontologies in the IoT-domain. *ArXiv*.
- 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>
- Barrachina, J., Garrido, P., Fogue, M., Martinez, F. J., Cano, J.-C., Calafate, C. T., & Manzoni, P. (2012). VEACON: A Vehicular Accident Ontology designed to improve safety on the roads. *Journal of Network and Computer Applications*, *35*(6), 1891–1900. <https://doi.org/10.1016/j.jnca.2012.07.013>
- Beil, C., Ruhdorfer, R., Coduro, T., & Kolbe, T. H. (2020). Detailed Streetspace Modelling for Multiple Applications: Discussions on the Proposed CityGML 3.0 Transportation Model. *ISPRS International Journal of Geo-Information*, *9*(10). <https://doi.org/10.3390/ijgi9100603>
- 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
- Bermejo, A. J., Villadangos, J. E., Astrain, J. J., Córdoba, A., Azpilicueta, L., Gárate, U., & Falcone, F. (2014). Ontology Based Road Traffic Management in Emergency

- Situations. *Ad Hoc Sens. Wirel. Networks*.
- Bermudez-Edo, M., Elsaleh, T., Barnaghi, P., & Taylor, K. (2015). *IoT-Lite Ontology*. (March 2017). Retrieved from <http://www.w3.org/Submission/iot-lite/%0A>
- BJSS. (2020). *Ontology Visualisation Proof of Value*. Retrieved from <https://www.bjss.com/highways-england-progresses-its-journey-to-a-data-driven-organisation/>
- Blakey, J. (2020). The politics of scale through Rancière. *Progress in Human Geography*. <https://doi.org/10.1177/0309132520944487>
- Bolton, A., Enzer, M., & Schooling, J. (2018). The Gemini Principles. In *University of Cambridge, UK 2018*. <https://doi.org/10.17863/CAM.32260>
- Brank, J., Grobelnik, M., & Mladenić, D. (2005). A survey of ontology evaluation techniques. In *Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2005)*. Retrieved from https://www.researchgate.net/publication/228857266_A_survey_of_ontology_evaluation_techniques
- Bricker, P. (2016). Ontological Commitment. In *Stanford Encyclopedia of Philosophy*. Retrieved from <https://plato.stanford.edu/entries/ontological-commitment/#QuiCriPre>
- Brunner, S., Kucera, M., & Waas, T. (2017). Ontologies used in robotics: A survey with an outlook for automated driving. *2017 IEEE International Conference on Vehicular Electronics and Safety (ICVES'17)*, 81–84. <https://doi.org/10.1109/ICVES.2017.7991905>
- 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>
- Carothers, G. (2014). RDF 1.1 N-Quads A line-based syntax for RDF datasets. Retrieved March 5, 2021, from <https://www.w3.org/TR/n-quads/>
- CDBB, U. of C. (2020). Top-Level Ontologies and Industry Data Models. Retrieved February 5, 2021, from <https://www.cdbb.cam.ac.uk/news/publication-top-level-ontologies-and-industry-data-models>
- 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>
- Centre-for-Argument-Technology. (2014). ARG-tech. Retrieved February 21, 2021, from Argument Mining website: <https://arg-tech.org/index.php/projects/argument-mining/>
- 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>
- CORDIS. (2020). European Science Vocabulary (EuroSciVoc). Retrieved March 5, 2021, from <https://data.europa.eu/euodp/en/data/dataset/euroscivoc-the-european-science-vocabulary>
- 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>
- Cummings, J., & Stacey, D. (2017). DCO: A Mid Level Generic Data Collection Ontology. *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KEOD 2017)*, 175–182. <https://doi.org/10.5220/0006497501750182>

- Dadzie, A. S., & Pietriga, E. (2017). Visualisation of Linked Data - Reprise. *Semantic Web*, 8(1), 1–21. <https://doi.org/10.3233/SW-160249>
- Dardailler, D. (2012). Road Accident Ontology. Retrieved February 7, 2021, from <https://www.w3.org/2012/06/rao.html#owl>
- data.gov.uk. (2021). INSPIRE examples. Retrieved from https://guidance.data.gov.uk/publish_and_manage_data/harvest_or_add_data/inspire/#examples-inspire-data
- 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>
- DEFRA. (2015). INSPIRE. Retrieved March 4, 2021, from INSPIRE website: <https://defradigital.blog.gov.uk/2015/02/25/inspired/>
- 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>
- Dou, D., McDermott, D., & Qi, P. (2005). Ontology Translation by Ontology Merging and Automated Reasoning. *Ontologies for Agents: Theory and Experiences*, 73–94. https://doi.org/10.1007/3-7643-7361-x_4
- Dsposal. (2021). KnowWaste GovTech Catalyst. Retrieved March 2, 2021, from <https://dsposal.uk/articles/knowwaste-govtech-catalyst/>
- Echterhoff, J., & Portele, C. (2017). INSPIRE RDF vocabularies. Retrieved March 4, 2021, from <https://github.com/inspire-eu-rdf/inspire-rdf-vocabularies>
- 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](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>
- EuropeanCommission. (2007). INSPIRE directive. Retrieved March 4, 2021, from <https://inspire.ec.europa.eu/about-inspire/563>
- Fielding, R. T. (2000). *Architectural styles and the design of network-based software architectures*, PhD Thesis. University of California, Irvine. Retrieved from https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf
- 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
- Gao, G., Liu, Y. S., Wang, M., Gu, M., & Yong, J. H. (2015). A query expansion method for retrieving online BIM resources based on Industry Foundation Classes. *Automation in Construction*, 56, 14–25. <https://doi.org/10.1016/j.autcon.2015.04.006>
- Garijo, D., & Osorio, M. (2020). OBA: An Ontology-Based Framework for Creating REST APIs for Knowledge Graphs. *ArXiv:2007.09206*. Retrieved from <https://arxiv.org/abs/2007.09206>
- GDS. (2020). (Government Digital Service) API technical and data standards (v2 – 2019). Retrieved from <https://www.gov.uk/guidance/gds-api-technical-and-data-standards>
- 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.

- Gerber, D., Nguyen, B., & Gaetani, I. (2019). *Digital twin: towards a meaningful framework*. Retrieved from <https://www.arup.com/perspectives/publications/research/section/digital-twin-towards-a-meaningful-framework>
- Gilbert, T., James, P., Smith, L., Barr, S., & Morley, J. (2021). Topological integration of BIM and geospatial water utility networks across the building envelope. *Computers, Environment and Urban Systems*, 86. <https://doi.org/10.1016/j.compenvurbsys.2020.101570>
- Gilbert, T., Rösndorf, C., Plume, J., Simmons, S., Nisbet, N., Gruler, H.-C., ... van Berlo, L. (2021). *Built environment data standards and their integration: an analysis of IFC, CityGML and LandInfra*. Retrieved from https://portal.ogc.org/files/?artifact_id=96354
- Gómez-Pérez, A. (2004). Ontology Evaluation. In S Staab & R. Studer (Eds.), *Handbook on Ontologies International Handbooks on Information Systems*. Springer, Berlin, Heidelberg (pp. 251–273). Retrieved from https://doi.org/10.1007/978-3-540-24750-0_13
- Grau, B. C., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., & Sattler, U. (2008). OWL 2: The next step for OWL. *Journal of Web Semantics*, 6(4), 309–322. <https://doi.org/10.1016/j.websem.2008.05.001>
- 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>
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220. <https://doi.org/10.1006/knac.1993.1008>
- Guarino, N., & Giaretta, P. (1995). Ontologies and Knowledge Bases: Towards a Terminological Clarification. In *Towards Very Large Knowledge Bases* (pp. 25–32). Retrieved from <http://www.loa.istc.cnr.it/old/Papers/KBKS95.pdf>
- Guizzardi, Giancarlo, & Wagner, G. (2004). A Unified Foundational Ontology and some Applications of it in Business Modeling. *CAiSE Workshops (3)*. Retrieved from <http://kischit.dyndns.org/sandbox/groups/tec/wiki/d694e/attachments/9cd96/UnifiedFoundationOntologies.pdf>
- Guizzardi, Giancarlo, 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.
- Hales, S. D., & Johnson, T. A. (2003). Endurantism, perdurantism and special relativity. *Philosophical Quarterly*, 53(213), 524–539. <https://doi.org/10.1111/1467-9213.00329>
- Harrison, W. (2006). Eating your own dog food. *IEEE Software*, 23(3), 5–7. <https://doi.org/10.1109/MS.2006.72>
- Hayes, S. (2021). DT Toolkit: Making the business case for a DT. Retrieved March 10, 2021, from <https://www.cdbb.cam.ac.uk/news/dt-toolkit-making-business-case-dt>
- Heiler, S. (1995). Semantic Interoperability. *ACM Computing Surveys*, 27(2), 271–273. <https://doi.org/10.2307/417557>
- Hetherington, J., & West, M. (2020). *The pathway towards an Information Management Framework*. Retrieved from <https://www.repository.cam.ac.uk/handle/1810/305579>
- 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>
- HM Infrastructure and Projects Authority. (2016). *National Infrastructure Delivery Plan 2016-2021*. Retrieved from <https://www.gov.uk/government/publications/national-infrastructure-delivery-plan-2016-to-2021>
- Hoehndorf, R., Dumontier, M., Oellrich, A., Rebholz-Schuhmann, D., Schofield, P. N., & Gkoutos, G. V. (2011). Interoperability between biomedical ontologies through relation expansion, upper-level ontologies and automatic reasoning. *PLoS ONE*, 6(7), 1–9. <https://doi.org/10.1371/journal.pone.0022006>
- Hogan, A., Harth, A., Passant, A., Decker, S., & Polleres, A. (2010). Weaving the pedantic Web. *LDOW2010, April 27, Raleigh, USA*. Retrieved from <http://ceur-ws.org/Vol->

- 628/ldow2010_paper04.pdf
- Hor, A.-H. (2015). *A semantic Web platform for BIM-GIS integration*.
<https://doi.org/10.13140/RG.2.1.4176.6643>
- Industrial Ontologies Foundry. (2021). IOF charter. Retrieved March 28, 2021, from
<https://www.industrialontologies.org/iof-charter/>
- Industrial Ontologies Foundry (IOF). (2021). Retrieved February 22, 2021, from
<https://www.industrialontologies.org/>
- IRIS. (2021). IRIS reasoner. Retrieved from <http://www.iris-reasoner.org/>
- ISO/IEC PRF 21838-2.2. (2021). Retrieved February 22, 2021, from
<https://www.iso.org/standard/74572.html>
- ISOTC22/SC32. *ISO 26262-1:2018 Road vehicles — Functional safety — Part 1: Vocabulary.*, (2014).
- 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](https://doi.org/10.1016/S0962-6298(97)00049-8)
- Judge, R., & Elahi, S. (2021). *Foresight review of the future of regulatory systems*. Retrieved from <https://www.lf-foundation.org.uk/en/publications/download-foresight-review-on-regulation/>
- 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*. 2518.
- Kavouras, M. (2005). A unified ontological framework for semantic integration. In P. Agouris & A. Croitoru (Eds.), *Next Generation Geospatial Information From Digital Image Analysis to Spatiotemporal Databases*. Retrieved from <https://b-ok.cc/book/951302/5da75e?redirect=147122266>
- Kellogg, G., Champin, P.-A., & Longley, D. (2020). JSON-LD 1.1: A JSON-based Serialization for Linked Data. Retrieved March 5, 2021, from
<https://www.w3.org/TR/json-ld11/>.
- Kharlamov, E., Skjaveland, M., Hovland, D., Mailis, T., Jimenez-Ruiz, E., Xiao, G., ... Waaler, A. (2019). Finding Data Should be Easier than Finding Oil. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, (December), 1747–1756.
<https://doi.org/10.1109/BigData.2018.8622035>
- Klímeck, J., Škoda, P., & Nečaský, M. (2019). Survey of tools for linked data consumption. *Semantic Web*, 10(4), 665–720. <https://doi.org/10.3233/SW-180316>
- 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>
- Lanthaler, M., & Gütl, C. (2012). On using JSON-LD to create evolvable RESTful services. *WS-REST 2012, Lyon, France*, 25–32. <https://doi.org/10.1145/2307819.2307827>
- 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>.
- Lee, Y. C., Eastman, C. M., & Solihin, W. (2016). An ontology-based approach for developing data exchange requirements and model views of building information modeling. *Advanced Engineering Informatics*, 30(3), 354–367.
<https://doi.org/10.1016/j.aei.2016.04.008>
- 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>
- LIU, L., & ÖZSU, M. T. (2009). Ontology. In *Encyclopedia of Database Systems*. Retrieved

- from https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-39940-9_1318
- Lorenz, B. (2005). *A1-D4 Ontology of Transportation Networks*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=075D8BCB207E43921DE630CBD832546E?doi=10.1.1.133.2953&rep=rep1&type=pdf>
- Lorenz, B., Ohlbach, H. J., & Yang, L. (2005). *Ontology of Transportation Networks*. Reasoning on the Web with Rules and Semantics.
- Lowe, E. J. (1995). Ontology. In *The Oxford Companion to Philosophy*. Oxford University Press (OUP).
- 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.
- Maedche, A., Motik, B., Stojanovic, L., Studer, R., & Volz, R. (2002). *Managing Multiple Ontologies and Ontology Evolution in Ontologging*. 51–63. https://doi.org/10.1007/978-0-387-35602-0_6
- Marinakakis, 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>
- Miles, A., & Bechhofer, S. (2009). SKOS Simple Knowledge Organization System Reference. Retrieved March 5, 2021, from <https://www.w3.org/TR/skos-reference/>
- Miller, D. (1991). Logic, Higher-order. In *Encyclopedia of Artificial Intelligence: Second Edition*. Retrieved from <http://www.lix.polytechnique.fr/~dale/papers/encyclopedia.pdf>
- Miller, E. (1998). An Introduction to the Resource Description Framework. *Bulletin of the American Society for Information Science and Technology*, 25(1), 15–19. <https://doi.org/https://doi.org/10.1002/bult.105>
- 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>
- Moreau, L., & Missier, P. (2013). PROV-DM: The PROV data model. Retrieved February 21, 2021, from <https://www.w3.org/TR/prov-dm/>
- Moreira, A., Araújo, J., & Brito, I. (2002). Crosscutting quality attributes for requirements engineering. *ACM International Conference Proceeding Series*, 27, 167–174. <https://doi.org/10.1145/568760.568790>
- Nevile, C. M., & Brickley, D. (2021). HTML Microdata. Retrieved March 5, 2021, from <https://www.w3.org/TR/microdata/>
- Niles, I., & Terry, A. (2004). The MILO: A general-purpose, mid-level ontology. *Proceedings of the International Conference on Information and Knowledge Engineering. IKE'04*. Retrieved from </paper/The-MILO%3A-A-General-purpose%2C-Mid-level-Ontology-Niles-Terry/cf7795baa3d976814268c132103034ad4d0e5bd1>
- Nisbet, N., & Liebich, T. (2007). *ifcXML Implementation Guide International Alliance for Interoperability Modelling Support Group: Herfordshire, UK*. Retrieved from <https://technical.buildingsmart.org/wp-content/uploads/2018/05/ifcXML-Implementation-Guide-v2-0.pdf>
- Noardo, F., Ellul, C., Harrie, L., Overland, I., Shariat, M., Arroyo Ohori, K., & Stoter, J. (2019). Opportunities and challenges for GeoBIM in Europe: developing a building permits use-case to raise awareness and examine technical interoperability challenges. *Journal of Spatial Science*, 65(2), 209–233. <https://doi.org/10.1080/14498596.2019.1627253>
- OGC. (2021). Why is the OGC involved in Sensor Webs? Retrieved from Open Geospatial Commission website: <https://www.ogc.org/domain/swe>
- OneM2M. (2021). Ontology for M2M. Retrieved February 22, 2021, from <https://www.onem2m.org/technical/onem2m-ontologies>

- Onerhime, E. (2020). *OpenDataManchester/KnoWaste*. Retrieved from <https://github.com/OpenDataManchester/KnoWaste>
- 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
- Partridge, C., Mitchell, A., Loneragan, M., Atkinson, H., De Cesare, S., & Khan, M. (2019). Coordinate systems: Level ascending ontological options. *Proceedings - 2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems Companion, MODELS-C 2019*, 78–87. <https://doi.org/10.1109/MODELS-C.2019.00017>
- Paulheim, H. (2015). What the adoption of schema.org tells about linked open data. *CEUR Workshop Proceedings*, 1362, 85–90. Retrieved from http://ceur-ws.org/Vol-1362/PROFILES2015_paper6.pdf
- Pease, A., Niles, I., & Li, J. (2002). The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. In *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*, 2002.
- 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
- Prud'hommeaux, E., & Carothers, G. (2014). RDF 1.1 Turtle Terse RDF Triple Language. Retrieved March 5, 2021, from <https://www.w3.org/TR/turtle/>
- Prud'hommeaux, E., & Seaborne, A. (2008). SPARQL Query Language for RDF. Retrieved March 5, 2021, from <https://www.w3.org/TR/rdf-sparql-query/>
- 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>
- Quine, W. V. O. (1980). *From a Logical Point of View: Nine Logico-Philosophical Essays* (Second Rev). Harvard University Press, reprint from Quine, W. V. (1948). "On What There Is". *Review of Metaphysics*. 2: 21–38.
- 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/>
- Rudnicki, R. (2019). *An Overview of the Common Core Ontologies*.
- Rudnicki, R., Smith, B., Malyuta, T., & Mandrick, C. O. L. W. (2016). *Best Practices of Ontology Development*. Retrieved from https://www.nist.gov/system/files/documents/2019/05/30/nist-ai-rfi-cubrc_inc_002.pdf

- Russom, M. B. (2021). Smart Cities Ontology for Digital Twins. Retrieved March 9, 2021, from <https://techcommunity.microsoft.com/t5/internet-of-things/smart-cities-ontology-for-digital-twins/ba-p/2166585>
- 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>
- Schmidt, D. (2020). Foundational Ontologies meet Ontology Matching : A Survey. *Semantic Web*, 0(0). Retrieved from <http://kyoto-project.eu/xmlgroup.iit.cnr.it/kyoto/index.html>
- Simons, P. M. (2015, January). Ontology. *Encyclopedia Britannica*. Retrieved from <https://www.britannica.com/topic/ontology-metaphysics>
- Singh, S., Shehab, E., Higgins, N., Fowler, K., Reynolds, D., Erkoyuncu, J. A., & Gadd, P. (2020). Data management for developing digital twin ontology model. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 095440542097811. <https://doi.org/10.1177/0954405420978117>
- Sirin, E. (2010). Data validation with OWL integrity constraints. In Hitzler, P., Lukasiewicz, T. (Eds) *Web Reasoning and Rule Systems. RR 2010 Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6333 LNCS, 18–22. https://doi.org/10.1007/978-3-642-15918-3_2
- Smith, B., & Ceusters, W. (2010). Ontological realism: A methodology for coordinated evolution of scientific ontologies. *Applied Ontology*, 5(3–4), 139–188. <https://doi.org/10.3233/AO-2010-0079>
- Stanford University. (2021). Protégé. Retrieved February 22, 2021, from <https://protege.stanford.edu/>
- Starks, G., Cheetham, M., & Patchay, J. (2020). *Enabling secure and scalable non-financial reporting and data flows (ICE Breaker 1 Data Infrastructure)*. Retrieved from <https://icebreakerone.org/report-nfdf/>
- 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>
- STFC. (2021). Data and Analytics Facility for National Infrastructure (DAFNI). Retrieved March 5, 2021, from <https://dafni.ac.uk/>
- Stojanovic, L., Maedche, A., Motik, B., & Nenad, S. (2002). User-Driven Ontology Evolution Management. *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web, 13th Int'l Conf EKAW*, 285–300. Retrieved from https://link.springer.com/chapter/10.1007/3-540-45810-7_27
- Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge Engineering: Principles and methods. *Data and Knowledge Engineering*, 25(1–2), 161–197. [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6)
- SWDWG. (2012). (Semantic Web Deployment Working Group) Simple Knowledge Organization System (SKOS). Retrieved from <https://www.w3.org/2004/02/skos/>
- Szilagyi, I., & Wira, P. (2016). Ontologies and Semantic Web for the Internet of Things - a survey. *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, 6949–6954. <https://doi.org/10.1109/IECON.2016.7793744>
- Taylor, M. (2017). Crossrail project: Building a virtual version of London's Elizabeth line. *Proceedings of the Institution of Civil Engineers: Civil Engineering*, 170(6), 56–63. <https://doi.org/10.1680/jcien.17.00022>
- TNO. (2015). Smart Appliances Project. Retrieved March 2, 2021, from <https://sites.google.com/site/smartappliancesproject/home>
- Togias, K., & Kameas, A. (2012). An ontology-based representation of the twitter REST API. *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*,

- Athens, Greece, 7-9 Nov, 998–1003. <https://doi.org/10.1109/ICTAI.2012.85>
- Tomai, E., & Prastacos, P. (2006). A framework for intensional and extensional integration of geographic ontologies. *Proceedings 2006 - The 9th AGILE International Conference on Geographic Information Science: "Shaping the Future of Geographic Information Science in Europe", AGILE 2006*, (April), 220–227.
- Topquadrant. (2021). Virtual Construction (V-Con) and TopBraid CDE. Retrieved from <https://www.topquadrant.com/vcon/>
- Trokanas, N., Cecelja, F., & Raafat, T. (2014). Towards a Re-Usable Ontology for Waste Processing. *Computer Aided Chemical Engineering*, 33, 841–846. <https://doi.org/10.1016/B978-0-444-63456-6.50141-1>
- 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>
- Tschirner, S., Scherp, A., & Staab, S. (2011). *Semantic access to INSPIRE: How to publish and query advanced GML data, Terra Cognita 2011, 23 October. Bonn, Germany.*
- 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>
- Verborgh, R., Sande, M. Vander, Hartig, O., Van Herwegen, J., De Vocht, L., De Meester, B., ... Colpaert, P. (2016). Triple Pattern Fragments: A low-cost knowledge graph interface for the Web. *Journal of Web Semantics*, 37–38, 184–206. <https://doi.org/10.1016/j.websem.2016.03.003>
- Verdonck, M., Gailly, F., & de Cesare, S. (2020). Comprehending 3D and 4D ontology-driven conceptual models: An empirical study. *Information Systems*, 93, 101568. <https://doi.org/10.1016/j.is.2020.101568>
- Vilgertshofer, S., Amann, J., Willenborg, B., Borrmann, A., & Kolbe, T. H. (2017). Linking BIM and GIS Models in Infrastructure by Example of IFC and CityGML. *ASCE International Workshop on Computing in Civil Engineering 25 June, Seattle Washington*, 133–140. <https://doi.org/10.1061/9780784480823.017>
- 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
- Vrandečić, D. (2009). Ontology Evaluation. In Steffen Staab & R. Studer (Eds.), *Handbook on Ontologies. International Handbooks on Information Systems, Springer, Berlin, Heidelberg* (pp. 293–313). Retrieved from https://link.springer.com/chapter/10.1007/978-3-540-92673-3_13
- W3C. (2020). Data Catalog Vocabulary (DCAT) - Version 2. Retrieved March 5, 2021, from <https://www.w3.org/TR/vocab-dcat-2/>
- Wagner, B., & Fain, N. (2018). Regulatory influences on innovation in the public sector: the role of regulatory regimes. *Public Management Review*, 20(8), 1205–1227. <https://doi.org/10.1080/14719037.2017.1350282>
- Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation Schemes*.

- <https://doi.org/doi:10.1017/CBO9780511802034>
- 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>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 1–9. <https://doi.org/10.1038/sdata.2016.18>
- Wright, A., Andrews, H., & Button, B. (eds). (2021). JSON Schema: A Media Type for Describing JSON Documents. Retrieved from <https://json-schema.org/draft/2020-12/json-schema-core.html>
- 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>
- Yagisawa, T. (2020). Possible Objects. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Summer 2020)*. Retrieved from <https://plato.stanford.edu/archives/sum2020/entries/possible-objects/>
- 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>
- Žáček, M. (2017). Ontology or formal ontology. *AIP Conference Proceedings*, 1863(July), 1–5. <https://doi.org/10.1063/1.4992234>
- 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
- Zittrain, J. L. (2009). The future of the internet - And how to stop it. In *The Future of the Internet - And How to Stop It*.