A SEMANTIC COMMON MODEL FOR PRODUCT DATA IN THE WATER INDUSTRY

Yasir Alani, PhD student
Teesside University
Y.Alani@tees.ac.uk

Nashwan Dawood, PhD, Professor
Teesside University
N.N.Dawood@tees.ac.uk

João Patacas, PhD, Research Associate
Teesside University
J.Patacas@tees.ac.uk

Sergio Rodriguez, PhD, Senior Lecturer
Teesside University
S.Rodriguez@tees.ac.uk

Huda Dawood, PhD, Senior Research Associate
Teesside University
H.Dawood@tees.ac.uk

SUMMARY: The issue of interoperability in the Architecture, Engineering, and Construction (AEC) industry represents a challenge on a scale that spans across the project life cycle. This is predominant in the infrastructure sector that usually comprises a more versatile Operations and Maintenance (O&M) phase in comparison with the buildings sector. To this end, an important stage in the information life cycle is the asset information capture and validation during product procurement at the O&M phase. The water industry in the United Kingdom relies on Product Data Templates (PDTs) to fulfil such task, which is usually an error prone manual process. This paper presents an ongoing research, which investigates the application of Semantic Web Technologies (SWT) for improving product data exchange during product procurement at the O&M phase for the water industry in the United Kingdom (UK). Therefore, focus group sessions with industry experts were held to discuss current inefficiencies and solution requirements. Based on these results, a semantic common model named Asset Specification Ontology (ASO) was developed to capture and validate asset information during product procurement at the O&M phase. The common model (ontology) is based on available technologies, namely Web Ontology Language (OWL) and Shapes Constraint Language (SHACL). This gives the advantage of semantically rich data which can be linked and queried in a meaningful way to facilitate the exchange and validation of water industry assets' data. The uniqueness of this paper is manifested in the issue it tackles, as efficient product procurement, and hence, data exchange in the water industry is an industrial challenge that is seldom researched.

Results from the focus group sessions showed that information exchange within the UK water industry is impeded due to the lack of structured and semantic data. However, for a robust semantic interoperability, there needs to be a robust semantic data infrastructure, which would require semantic mappings from standards to product properties, from standards to other standards, and from standards to dictionaries. These conclusions were further supported by the common model, which was created from existing schemas, standards, and dictionaries. Generally, this paper recommends a common model/product library for phase-specific product data exchange in the water industry.

KEYWORDS: Product data exchange; Product Data Templates; PDT; Water industry; Semantic Web Technologies; data validation; Operations and Maintenance

1 INTRODUCTION

Civil engineering projects are usually complex in terms of information management and flow. The large number of stakeholders involved in such projects makes information management a non-trivial task, particularly for infrastructure domains, where the adoption of Building Information Modelling (BIM) is lacking behind (Bradley et al., 2016). Cheng et al. (2016) divided the infrastructure assets into five domains, namely, 1) transport, 2) energy, 3) utility, 4) environmental, and 5) recreational facilities. As this research focuses on the water sector within the utility infrastructure domain in the United Kingdom, two of the major reasons for poor utility management is the absence of a digitalised information model with a comprehensive database, and consequently, a poor utility information representation (Wang et al., 2019) and integration (Bradley et al., 2016). This information is predominantly non-graphical data (Bradley et al., 2016). One example of non-graphical data is product specifications, necessary for the Operations and Maintenance (O&M) phase. Precisely, the acquisition and validation of product data within the life cycle of information flow. This is a prominent problem even in the building domain (discussed in sub-section 1.2) as Pauwels et al. (2017) regarded combining product manufacturer data with building data as a research challenge, as the acquisition and validation of product data is generally a manual process that involves static documents (Rasmussen et al., 2019). The most common product data capture methodology in the United Kingdom's (UK) water infrastructure sector is the Product Data Templates (PDTs) for capturing non-graphical data. PDTs are generally constructed using product standards and serve as a common framework for managing construction product data (CIBSE - Building Information Modelling, n.d.; Product Data Template, 2020). Nonetheless, the format of these templates remains unspecified as well as the methodology of their use for data exchange. Current practice adopted by water companies utilises PDTs in the XLS format, which makes data exchange and validation manual and lengthy as spreadsheets are less useful when an application requires highly interconnected information (Allemang and Hendler, 2011). Examples of the PDTs for the water industry in the UK can be found on British Water website (British Water, n.d)

This research aims to improve the process of product data specification comparison, capture, and validation for the water industry during procurement at the O&M phase. The O&M phase was chosen because it includes mature and specific product requirements, and thus, it is possible to conduct specification comparison and data validation. This paper 1) explores current inefficiencies and gathers the relevant requirements for solutions from industry practitioners in the water domain via focus group sessions, and consequently, 2) proposes a semantic common model to address these requirements. This common model presented in section 4 was developed using the Web Ontology Language (OWL) and the PDTs as an explicit resource to represent product information. The common model would form part of a common model library, which with the aid of Shapes Constraint Language (SHACL) would be used for product data specification comparison, capture, and validation. Therefore, throughout this paper, *data exchange* will refer to the procurement of assets based on asset specification requirements comparison and validation at the O&M phase. The subject of this study is the water industry in the UK. Sub-sections 1.1 and 1.2 briefly discuss Semantic Web Technologies (SWT) and explore their application in the field of product data modelling for AEC.

1.1 SWT in AEC

The most common semantic modelling language is the Resource Description Framework (RDF), which is a W3C recommendation (Cyganiak et al., 2014). A piece of information represented in RDF is known as a triple, which consists of the "subject-predicate-object" relationship, Fig. 1. The object of one triple would form the subject of another and so on, creating a chain of triples connected via predicates called object properties. Predicates that lead to data values are called data properties. Using such language, different knowledge domains can be represented as ontologies. The OWL and RDF Schema (RDFS) are ontology modelling languages built on top of RDF, that differ in their complexity, and consequently, inferencing capability. Gruber (1993) and Borst (1997) defined an ontology as an "explicit specification of a shared conceptualisation." Thus, an ontology is ought to be formal, sharable, extendable, and reusable. These characteristics have encouraged several research proposals to represent AEC domains using ontologies (Pauwels et al., 2017), confirming their usefulness in achieving interoperable data exchange. Perhaps the most prominent one is the ifcOWL ontology (Beetz et al., 2009; Pauwels and Terkaj, 2016). Pauwels et al. (2017) proposed modularising large ontologies to reduce their size and complexity, by splitting them into smaller ontologies, in which end users would only load the relevant modules they need. They claimed that ifcOWL did indeed provide the IFC schema with better interoperability, linking across domains, and logical inference and proofs. However, they (Pauwels et al., 2017) developed the claim that there is a trade-off between

semantic precision and computational efficiency. This trade-off results in reduction in expressive power for better reasoning efficiency.

FIG 1: RDF triple

Niknam and Karshenas (2017) presented a shared ontology approach for semantic BIM, where every domain develops its own ontology through extending a shared (top-level) ontology. Niknam and Karshenas (2017) argued that the AEC-FM industry lacks a standard ontology that can be used for converting BIM data to a semantic format, as expanding a single ontology (i.e. ifcOWL) to include the all relationships in the AEC-FM industry is unreasonable due to the fact that ontology mapping is a time consuming and error prone, manual process. Nonetheless, their work was aimed at the buildings sector. Kim et al. (2018) claimed that BIM information is usually not utilised well during the operational phase. So, they proposed linking BIM data to historical work records through a semantic web-based FM database. Kim et al. (2018) tackled the problem of storing, merging, and retrieving heterogenous information. They stated that BIM data are object-based whereas FM data are activity-based, and thus, they utilised SWT to mediate between the two modelling paradigms and reported benefits to object-based FM using SWT. However, Kim et al. (2018) claimed that to use their method, facility managers should be trained in both BIM and SWT. Tchouanguem Djuedja et al. (2021) proposed integrating LCA data with BIM software through RDF graphs. Tchouanguem Djuedja et al. (2021) stated that SWT remains a promising solution for data sharing and interoperability between heterogenous systems. Also, SWT have a feature of interest in the building energy domain. Researchers such as Howell et al. (2017); Yuce and Rezgui (2017); Corrado et at. (2015); Hippolyte et al. (2018) have employed ontologies for the semantic modelling of energy related data.

Pauwels et al. (2017) reviewed SWT in the AEC industry and concluded that SWT are an important addition to current technologies. However, they argued that the creation and maintenance of data links requires human intervention, resulting in interoperability issue on a finer scale, the data level. However, Bradley et al. (2016) proposed that knowledge graphs facilitated by SWT and Linked Data (LD) provide a schema independent model that enables a dynamic information definition and resource linking. They further concluded that knowledge graphs are the way forward in managing the vast amount of data in construction projects. Jaskó et al. (2020) concluded that in the future, formal models such as ontologies will play an essential role in Industry 4.0 systems as the demand for semantically rich information increases. Generally, one can conclude that there is ample evidence in the literature developing the claim that SWT enhance interoperability. Thus, SWT were used as technological underpinning for this research, which manifests in the capture and validation of product data for the water industry in the UK during product procurement at the O&M phase.

1.2   SWT for AEC product data

Several researchers have tackled the issue of building product data modelling and exchange using SWT. Beetz and de Vries (2009) stated that the lack of machine-readable building product specification forms an issue and proposed semantic products catalogues using SWT for the building domain. Abdul-Ghafour et al. (2007) proposed a Common Design Features Ontology (CDFO) to act as a mediator between CAD systems trying to exchange non-geometrical information. Abdul-Ghafour et al. (2007) concluded that the challenge lies in defining the mapping rules to convey semantic meaning between a CAD system and other systems. For instance, some concepts may not have semantic equivalences in other systems. Abdul-Ghafour et al. (2014) used Description Logic (DL) rules to achieve semantic mapping between ontologies to exchange CAD product data. Costa and Madrazo (2015) stated that the participation of manufacturers in the modelling process is usually impeded. Costa and Madrazo (2015) developed a BIM compatible online product catalogue for the building domain using SWT, as they claimed that BIM technologies usually do not provide links to product components. Costa and Madrazo (2015) concluded that SWT can help interlinking different domains, and hence, integrate data. However, they did not mention the process of data validation once it is captured. Gao et al. (2017) claimed that building product libraries often contain ambiguous BIM documents for product descriptions, which would require manual annotation to classify and tag

the information. Gao et al. (2017) created an ontology by converting the relevant IFC concepts, which was then used to automatically annotate/retrieve BIM documents. However, as the IFC is focused on the building domain, similar approaches for the infrastructure sector remain lacking. Bilal et al. (2017) developed an ontology, dubbed data-agnostic semantic repository, to describe building materials following specialised requirements such as Uniclass classification. The property-based Ontology for Property Management (OPM) was developed to model properties in the design environment (Rasmussen et al., 2019). Rasmussen et al. (2019) also proposed an API to query their ontology and retrieve design data. However, OPM can be considered as a top-level ontology for describing properties' metadata, rather than modelling products. Rasmussen et al. (2019) concluded that "SWT can be used to cope with the highly interrelated and rapidly changing design decisions when developing a construction project". Wagner and Rüppel (2019) presented the Building Product Ontology (BPO) for building façade components and discussed some relevant ontologies. However, these ontologies including BPO can be regarded as top-level alignment tools rather than detailed data models.

Similar approaches for the water industry are lacking, as research on utilising SWT in the infrastructure sector is lacking in general (Ren et al., 2019). Ren et al. (2019) created an ontological knowledge base for bridge maintenance and claimed that their ontology is used for asset management and automatic rule checking. Ren et al. (2019) also claimed that their ontology can help identify relevant material suppliers. However, suppliers' data must be stored in the ontological knowledgebase in the first place. El-Diraby and Osman (2011) suggested that informatics ontologies (as opposed to philosophical ontologies) are divided into upper, domain, and application ontologies and proposed a domain ontology for infrastructure products that is aimed at knowledge representation rather than inferencing or data validation. However, research on using SWT for solving the issue presented in this paper for the water domain - where online 3D product models are less common (Bradley et al., 2016) - remains lacking. Although, several standards exist that support the representation and exchange of product data such as ISO 10303, ISO 15926, ISO 23386, ISO 23387, and IEC 62264, applications of these standards within the context of this research were not observed in literature. Table 1 provides an overview of the papers that were investigated in this research. This is not an exhaustive review of the literature, however. This was used to support the requirements capture process (section 3) for a better understanding of the gaps in research. In this regard, it is apparent that there is a gap in research with respect to using ontologies to model and capture infrastructure assets data, as previous work has primarily focused on developing top-level ontologies for the building domain, which has significant differences to the infrastructure domain, as the latter has a different work breakdown structure that requires a more mature asset management process (Bradley et al., 2016).

1.3 Aims and objectives

This paper addresses the current inefficiencies of product data exchange during product procurement at the O&M phase in the UK water sector. These inefficiencies are further explored in section 3, based on which, an ontological common model was proposed in section 4. The research questions in this paper can be summarised as follows:

- What are the current inefficiencies in the UK water industry in terms of asset data acquisition during product procurement?
- How can these inefficiencies be addressed using SWT?

Therefore, this paper presents two aims that contribute to the existing body of knowledge:

- Explore the current problems and requirements for product data exchange during product procurement in the water industry using focus groups with key stakeholders.
- Propose a domain-specific ontological common model, as part of a common model library, to address the problems and meet requirements.

The novelty of such common model lays in its abstraction principles of representing a product with a high level of detail that is suitable for the water industry in the UK; this combined with the inference and validation capabilities of the common model, high quality detailed product information is expected during product procurement.

In the remainder of this paper, section two presents the research methods. Section three presents the focus group sessions. Section four introduces part of the ontological common model development process. Section five gives a brief overview of the data validation method used with the common model. Section six describes the use of the common model within the context of international standards. Section seven provides discussion on the results from

focus group sessions (requirements capture) and the developed common model. Section eight provides conclusions. Finally, Section nine summarises the limitations and future work plans.

TABLE 1: Papers that discussed semantic product data in construction

Paper	Domain of ontology for capturing asset data						Ontology for Product data	Ontology for IR
	Buildings	Infrastructure				Other (general)		
		Water	Oil and gas	Transport	Energy			
Abdul-Ghafour et al. (2007)						x	x	
Li and Ramani (2007)						x		x
Beetz and Vries (2009)	x						x	
Böhms et al. (2009)	x						x	
El-Diraby and Osman (2011)						x	x	
Chaprala et al. (2013)	x						x	
Hahm et al. (2014)						x		x
Corrado et al. (2015)					x			
Costa and Madrazo (2015)	x						x	
Gao et al. (2015)	x							x
Gao et al. (2017)	x							x
Akinyemi et al. (2020)		x					x	
Kim et al. (2018)	x							
Ren et al. (2019)				x			x	
Wagner and Rüppel (2019)	x						x	
Wu et al. (2019)	x						x	

2   RESEARCH METHODS

The gap in the literature has been identified as the lack of semantic modelling of product information in infrastructure projects in general, and the water industry in particular. This has generated motivation to collect requirements from practitioners within the water industry, and so, focus group sessions were held to discuss current problems pertinent to product data flow and set the requirements for solutions. Therefore, the focus group research method, that is presented in section 3, was used to answer the first research question in this paper. Then, based on the results that emerged from focus groups analysis, an ontological common model was developed using the NeOn methodology (Suárez-Figueroa et al., 2012), which is presented in section 4. The ontological common model alongside the SHACL validation language (presented in section 5), which emerged as a requirement from the focus groups, tackled the second research question in this paper. Fig. 2 shows the research process.
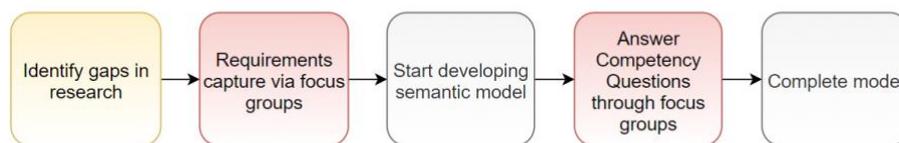

FIG 2: Research methods

3   REQUIREMENTS CAPTURE

For capturing the needs of the water industry in the UK within the context of the issues presented in section 1 - i.e. product procurement and data exchange, a pilot interview was conducted with an asset data manager for a water company in the UK; participant 1 in Table 2. Then, iterative focus group sessions were held online with five

experts in the industry. The goal of these focus group sessions was finding answers to the first research question presented in sub-section 1.3. The focus group method for data collection was selected, due to the small number of industry experts on the matter, and the purpose of generating interaction within the respondents (Gray, 2018). According to Seal et al. (1998), interviews are suitable for an in-depth identification of an individual's opinion, whilst focus groups are better at generating discussions about shared and unshared experiences. Focus groups are more capable of filtering out inaccurate information, as they provide a wider source of expertise and lead to views that are dominant among participants (Elsawah et al., 2017). Powell, Single and Lloyd (1996) defined a focus group as "a group of individuals selected and assembled by researchers to discuss and comment on, from personal experience, the topic that is the subject of the research". The focus groups were guided by few questions that stirred the discussion among participants, all within the context of the first research question, i.e. what are the current inefficiencies in the UK water industry in terms of asset data acquisition for product procurement? These questions are presented as competency questions for the ontology development in Table 4. Description of the participants is presented in Table 2. Additionally, meeting the following conditions helped strengthen the requirements capture process (Naoum, 2014):

- Respondents were homogenous and of related disciplines.
- Discussions were within the scope of the respondent's knowledge.
- Interpersonal contact was provided through a presentation about common models, and explanation of the objectives.
- The discussions required detailed answers

TABLE 2: Focus groups participants

Participant	Role	Organisation
1	Head of data Management	Water Company
2	Global BIM manager	Water technology company (manufacturer)
	Chair	Water asset management company (BIM)
3	Asset data team manager in	Water company
	Working group chair	Water asset management company (BIM)
4	Project leader	CEN working group
	Member	ISO working group
5	Enterprise Data Architect	Water company
	Chair of task group	Water asset management company (BIM)
6	Consultant	Water industry's digital delivery

Qualitative analysis was used for generating a hypothesis based on the data collected and interpretation of that data (Pickard, 2013). Constant comparative analysis was used to analyse the data collected from the focus groups. This strategy is based on comparing pieces of data with each other to develop conceptualisations about the relations between them (Pickard, 2013). The word data in this context refers to any unit of analysis that is relevant – a word, sentence, or a paragraph. The first stage in the process of analysis, namely, "microanalysis", is open coding which involves examining the data line by line to identify categories. The second stage is axial coding to establish the links between categories and subcategories. The final stage is selective coding which is ought to refine the generated theory (Pickard, 2013). So, it is during the final stage when the grounded theory is developed. According to Corbin and Strauss (1998), the grounded theory should emerge from a central category that has appeared in all the other categories.

Table 3 shows the main categories and themes that emerged from the focus groups. Results suggested that there are indeed inefficiencies in terms of product data acquisition, validation, and structuring. One of the reasons for these inefficiencies is the lack of a unified classification system for the water industry in the UK. Although Uniclass aims to provide classification system, when respondents were shown some of the inconsistencies that arise when modelling Uniclass, they agreed that restructuring may be needed. In terms of data acquisition, respondents encouraged having a standardised common process to capture asset data. In this regard, the PDTs provide some consensus knowledge in the domain. However, issues of data quality, ownership, and semantics remain unsolved. Respondents agreed on that a solution for data exchange should be developed from existing schemas and incorporate data dictionaries and standards. Also, they highlighted the importance of clearly defining the context of data exchange, as the requirements for data exchange would differ during each phase of the project. Hence,

solutions that facilitate data exchange should be phase-specific yet must fit within the whole life cycle of information flow. This means that each phase of data exchange should align with the phases that precede and succeed it. Also, a solution should account for the needs of the entire industry, as well as accommodate the similarities with other industries. Results also suggested that although standards are important in aiding semantics, they can be ambiguous. Respondents highlighted the importance of data validation via setting constraints on the data. Respondents agreed on that the process of product data exchange should consist of three stages; 1) a data dictionary such as the PDTs to identify the requirements, 2) a message server between the parties involved, and 3) a common model server to facilitate data exchange and validation. Generally, respondents encouraged creating a detailed asset specification model that can be used for procurement and data validation. In summary, the theory that emerged from the focus groups analysis can be phrased as follows; Information exchange in the water industry in the UK is impeded due to the lack of structured and semantic data.

From these results, one can conclude that a solution is ought to provide a holistic, standardised, and web-based information structure that enables a relatively easy understanding, querying, mapping, and validation of product data for the water industry in the UK. Thus, considering these results, the recent research on utilising SWT in AEC, and other published literature - such as ISO standards and RIBA (Royal Institute of British Architects) Plan of Work, an ontological common model was developed.

TABLE 3: Focus groups results after Constant comparative analysis

Topic	Categories	Themes
Classification	Current inefficiencies	• Information is classified differently within the same company • Requirements differ between water companies • Uniclass has shortcomings • There is no method of data validation • There are no thresholds on the data
	Importance of classification	• The owner/operator would get what they need from day one • Water companies should have connected AIMs via classification • Understanding data requirements is key to decision making
	PDTs	• PDTs are a starting point • PDTs are created through an agreed process • PDTs are for maintenance phase to aid communication between capital delivery and supply chain
	Reference data library	• Reference data libraries lack semantic relations
	Uniclass	• Uniclass is not detailed enough • Uniclass is only partially adopted • The structure of Uniclass can be contradictory
Data	Data quality	• How to validate data input as well as output? • Data exchange should result in 80% accuracy, at least • Constraint languages are important • How to validate the common model?
	Data ownership	• Data ownership is an issue
	Semantic meaning	• Ontologies are important for semantic meaning • Standards could aid semantic meaning
	Common process	• There should be a common process for data exchange • Data should be captured via PDTs and managed via a central warehouse
Digital twin	National Digital twin	• There should be modularity in a national digital twin
Data lifecycle	Data lifecycle requirements	• Data requirements are not clear at the beginning
Mapping	Mapping classifications	• Classifications should be converted into properties, then into connected properties, then into dictionaries, and then connected to standards
	Mapping between systems	• Mapping classifications is key to mapping between systems
Similarities	Similarities within industry	• Water companies share the same Outcome Delivery Incentives, assets, and top-level classifications
	Similarities with other industries	• The common model must account for the fact that some assets are used by more than one industry

Topic	Categories	Themes
Standards and dictionaries	Importance of standards	• Standards supporting the process of information exchange are being developed • Some efforts standardise asset data by breaking them into functional modules • Standards are important for their consensus and supporting semantics
	Lack of standards	• PDTs rely on an oil and gas standard • There are differing naming conventions between standards • This common model is relevant from a standardisation point of view
The common model	Ontology definition	• We must use existing schemas for the common model • Reference external documents • Object models can be used as data catalogues or collaborative platforms • A common model using OWL and SHACL is useful, as there is no ISO implementation • We should define the necessary steps by water companies to utilise such common model
	Data validation	• Constraints are important • SHACL can be used at different phases • There should be a data exchange and validation use case between two actors • The use case should simulate the process negotiation • Value ranges are important • Some water companies do not specify thresholds for their asset data (ranges) • The common model can be validated by consulting more experts after the use case
Need for use case	Important steps in the use case	• The use case must be specific to a certain phase • Negotiations start with the process, identifying the actors, identifying information exchange requirements • Once you have the business standards and the business process, then you have a PDT • There should be 3 phases to a process; 1) data dictionary, 2) message server, and 3) a model server • We must define the context of the use case (process)
	Owner's requirements	• Owner's requirements are important • Owner's requirements (water companies) vary across the UK

4   ASSET SPECIFICATION ONTOLOGY (ASO) DEVELOPMENT

From the results of the focus group sessions, there needs to be a standardised, web-based information exchange of product data between suppliers and designers. Semantic Web Technologies were chosen to provide the necessary tools for this research, and hence, the second aim of this paper was to develop an ontological common model named ASO. According to Suárez-Figueroa et al. (2012), the number of published ontologies is rapidly increasing on the web, and therefore, "in this context ontology development can be then characterised as the construction of a network of ontologies, where the different resources may be managed by different people, possibly in different organizations". After consulting the literature on ontology building methods such as Uschold and King (1995), Grüninger and Fox (1995), and Gómez-Pérez et al. (2004). The NeOn methodology was utilised in creating the ontological common model (Suárez-Figueroa et al., 2012) due to it being well recognised by researchers in the AEC industry such as Li et al. (2019) and Kuster et al. (2020), and its scenario-based approach. It identifies nine scenarios that may arise during ontology development. The common model development requirements – as emerged from the focus groups – match the second scenario, which is "reusing and reengineering non-ontological resources". The PDTs formed the main non-ontological resource because they have domain consensus and incorporate domain standards. Some of the development process of the common model is presented in this section. This common model is a work in progress; more data will be collected upon future case studies, to verify or refute its effectiveness in bringing efficiency to the product procurement process. Also, future work will focus on modularising this ontology to create a network of ontological models describing the water industry in the UK to improve efficiency.

The NeOn methodology specified four major steps in ontology development namely, specification, conceptualisation, formalisation, and implementation. The next sections present the ASO development according to these four steps.

4.1 Ontology Requirements Specification

The ontology specification is the highest level of abstraction and the starting point in ontology construction, which can be summarised in the Ontology Requirements Specification Document (ORSD) (Suárez-Figueroa et al., 2012). The ORSD specifies the requirements that the ontology developer must fulfil to start building the ontology. These requirements include understanding 1) the purpose, 2) the scope, 3) the implementation language, 4) the intended users, 5) the functional and non-functional requirements, and 6) the pre-glossary of terms of the ontology.

4.1.1 Purpose

As stated earlier, the purpose is to create a common model for water assets specifications using PDTs as the main schema source. This ontological common model is ought to form a part of a common model library , which would be used as a standardised tool during the data exchange process for water assets. Data exchange in the context of this paper refers to procurement of assets based on asset specification requirements comparison and validation at the O&M phase.

4.1.2 Scope

The scope of this common model will be limited to the data schema in the PDTs for the water domain, following ISO 23386 and ISO 23387 (British Standards Institution, 2020; British Standards Institution, 2020). The common model targets the O&M phase procurement specifically.

4.1.3 Implementation language

OWL was chosen to be the language for the ontology. Ideally an ontology is designed with one of the four OWL 2 subsets in mind, namely DL, EL, RL, and QL. Each one of these subsets targets a specific functionality. The ASO model was designed using OWL 2 DL (Description Logic) to aid decidability and computational completeness. OWL 2 DL includes OWL 2 EL, OWL 2 RL, and OWL 2 QL.

4.1.4 Intended users

This ontology is intended to be used by designers when procuring assets from suppliers at the O&M phase. So, the intended users are designers and suppliers.

4.1.5 Ontology Requirements

Ontology requirements are divided into functional and non-functional requirements:

- *Non-functional requirement.* These requirements refer to the characteristics of qualities that the ontology is ought to have such as standards and naming conventions. For example, according to results from the focus group sessions, the Uniclass code for each asset should be included in the model. Also, the naming convention for properties should be as similar as possible to that found in the PDTs due to it being based on the buildingSMART Data Dictionary (bsDD).
  Functional requirements: Competency Questions (CQs). CQs are used to understand the high-level ontology development requirements, and hence, were used in this research to guid the focus group sessions in section 3. Table 4 presents the CQs alongside their answers. Answers to these CQs were formed based on the authors' understanding of the focus groups results and ontology development practices found in literature.

TABLE 4: Ontology Competency Questions

Competence question	Answer
CQ1. Why would one develop this model?	The main reason for developing an ontology is to aid solving the research problem/theory that emerged from the focus group sessions. The purpose of the ontology, therefore, is to act as a phase-specific, product information exchange tool used for product procurement.
CQ2. What domain will the model cover?	This ontology is ought to cover the asset specifications requirements when procuring assets at the O&M phase for the water industry in the UK. Hence, it will cover the water domain in the UK.

Competence question	Answer
CQ3. What are the expected outcomes/results of this model?	The ontological model is expected to improve data exchange during asset procurement. Improvements are expected in terms of data quality due to better reasoning and validation. Acquiring high quality data in less time has implications that could span across the entire lifecycle of an asset, as the information in the PDTs is currently used to manage the assets across different platforms.
CQ4. What types of data will the model represent?	The ontology, as the PDTs, is ought to represent static non-graphical asset manufacturer data. These data can be quantitative as well as qualitative.
CQ5. What are the inputs and outputs of this model?	The inputs will be asset specification data found in the PDTs, such as manufacturer, serial number, cost, etc. These data are static data that come from the manufacturers or suppliers. However, it is important that they could be linked to dynamic (functional) data in the future. The outputs include smarter web-based knowledgebase that allows querying good quality data, that can be validated.
CQ6. Considering the relatively large schema size in the PDTs, what is the most suitable way to convert this to OWL (one ontology or separate ontologies)?	The ontology should be developed with the "component" class constituting the core of the ontology. It is important that the ontology can be modularised at later stages to form part of a national digital twin.
CQ7. If this model was to cover asset specifications, then what is meant by an asset within this context?	An asset is a component that is, a physical entity and of value to the water company. Hence, it is a component that can be procured from a supplier. This definition agrees with ISO 19650-1 (British Standards Institution, 2018).
CQ8. How to represent a non-Uniclass component and a Uniclass component?	A Component is a Uniclass product if and only if it has a Uniclass code and certain necessary mandatory properties that are found in every PDT.
CQ9. How important is data validation?	Data validation is very critical. Currently, information in the PDTs is validated manually and without the ability to set thresholds on the data. Validation here means that data are present, in correct format, have the correct units, and are within a defined range (if applicable). More advanced validation is an advantage such as IF-THEN rules and mathematical equations check.
CQ10. In asset management, what level of detail is sufficient? In other words, would the PDTs be more beneficial if they were more detailed, or more abstract?	It was realised during the focus group sessions that the PDTs are more detailed than the current need of the water companies in the UK. However, the need for more asset data is expected to increase with time and technology. Therefore, the level of detail in the PDTs was to be deemed sufficient for the common model.

4.1.6 Pre-glossary of terms

The pre-glossary of terms is a list of terms that can be extracted from the ontology requirements identification process, which can be translated into classes, properties, and data types. These are the terms that appeared frequently in the competency questions, their answers, and named entities. However, this step is important if one was to create an ontology from scratch, which is not the case in this research. Therefore, the pre-glossary of terms was mainly extracted from the PDTs.

4.2 Conceptualisation

In this level of ontology development, the structure and components of the ontology are described. It is this step that introduces the main definitions and abstraction principles that constitute the ontology. The key concepts in this ontology were extracted from the PDTs following the industry's requirements/needs realised from the focus groups.12 PDTs were modelled (presented in table 5) in the ontology. Generally, one of the challenges was deciding whether to model something as a class instance or data value. Several aspects were taken into consideration when making this decision such as querying needs, schema population from other sources, model requirements, mapping requirements, and if the entity to be modelled had other dependencies, e.g. a material would have material grade. The ontology in this research includes the following key concepts:

1. A component can refer to an undividable physical asset or a physical asset that is made of a collection of smaller (undividable) components. Any asset mentioned in the PDT that may need maintenance and/or future replacement is a component. Every Uniclass preparatory product is component but not every component is a Uniclass preparatory product.

2. A component is a Uniclass product if and only if it has a specific list of restrictions. These restrictions are the necessary (mandatory) properties that all PDTs have in common such as Uniclass code, manufacturer, serial number, etc. Therefore, the class of Uniclass preparatory products is a defined class.

3. Based on the relationships in the PDTs, every subclass of UniclassPreparatoryProduct is defined by a list of class restrictions as a primitive class. These restrictions represent properties that are specific to an asset type. For example, there are specifications that are particular to a submersible pump only – in addition to the inherited mandatory properties.

4. Globally Unique Identifiers (GUID) of assets are important to water companies and recommended by ISO 12006-3 (British Standards Institution, 2016) and ISO 23386 (British Standards Institution, 2020). Every Uniclass preparatory product has a GUID serving as an asset identification code. The same concepts apply for an asset's serial number. However, GUIDs are for later use as they are generated internally by the water company when an asset is installed, while serial numbers are generated by the asset manufacturer.

5. Data properties were kept to a minimum. Only simple properties that had no other dependencies were modelled as data properties. For example, a property that only required a Boolean value without further restrictions is modelled as a data property.

6. Any measurable property associated with a component was modelled as a class and defined using the Quantity Kind definition borrowed from the Quantities, Units, Dimensions, and Types (QUDT) ontology (QUDT, 2021). Such Properties include overall height, overall length, overall width, fluid temperature, power rating, pressure, pump flow rate, voltage, etc.

7. Any property that requires property management to accommodate design changes by manufacturers was modelled as a class, to allow mapping to OPM (Rasmussen et al., 2019) for example. Hence, property management definition is aligned with the definitions of Rasmussen et al. (2019) and ISO 23386 (British Standards Institution, 2020). It is worth noting that OPM would require updating to be completely conformant with ISO 23386.

4.3 Formalisation and implementation

The ontology schema was developed from 12 PDTs listed in Table 5 as a Proof of Concept (PoC). Protégé 5.5 was initially used to create the ontology (protégé, 2020). The ontology was given the namespace *aso* (asset specification ontology) as initially published in (Alani et al., 2020). However, major revisions were carried out in this paper.

TABLE 5: The PDTs used to produce the ontology classes in Fig. 3

Product Name	Uniclass Code
Submersible pump unit	Pr_65_53_24_86
End suction pump (back pull-out type)	Pr_65_53_96_06
Progressing cavity pump	Pr_65_53_96_66
Vertically suspended bowl type pump	Pr_65_53_96_94
Electromagnetic flow meter	Pr_80_51_46_27
Instrumentation, automation, and control panel	Pr_75_50_18_41
Incoming supply control panel	Pr_75_50_18_40
Motor starters control panel	Pr_75_51_52_22
Pressure sensor	Pr_75_50_76_68
Ultrasonic sensor	Pr_75_50_47_92
Distribution board	Pr_60_70_22_22
Metal penstock	Pr_65_45_30_52

Regarding concept 1 from sub-section 4.2, a primitive class named aso:Component was created. The class aso:Component is an rdfs:subClassOf owl:Thing. The component class is at core of the ASO, as it includes all the relevant physical assets that are of value, as defined in the conceptualisation stage. The classes presented in Fig. 3 are examples of rdfs:subClassOf aso:Component that were extracted from the 12 PDTs. Some of these components have subclasses, which are also of type aso:Component through inheritance. For instance, class aso:Pipe, shown in Fig. 3, is an aso:Component and has further subclasses such as 1) aso:ColumnPipe which forms part of (aso:formsPartOf) aso:VerticallySuspendedBowlTypePump, and 2) aso:MeasuringPipe which forms part of

aso:ElectromagneticFlowMeter. Components are connected via the aso:hasComponent object property, which is defined Listing 1. The aso:hasComponent is declared as an owl:TransitiveProperty, the effect of which is demonstrated in Fig. 4.

```
        aso:hasComponent rdf:type owl:ObjectProperty,
            owl:TransitiveProperty ;
        rdfs:domain aso:Component ;
        rdfs:range aso:Component ;
        owl:inverseOf aso:formsPartOf .
```

LISTING 1

Regarding concepts 2 and 3, one can see from Fig. 3 that every UniclassPreparatoryProduct is a Component. However, a component is a UniclassPreparatoryProduct if it has certain properties (definitions) that are represented as restrictions in OWL. An example of these restrictions is demonstrated in Listing 2. This example means that any instance that has value for the property aso:hasSerialNumber from class aso:SerialNumber will be inferred as a UniclassPreparatoryProduct. An intersection of many similar restrictions make up the definition of a UniclassPreparatoryProduct.

Fig. 5 shows an example of how a UniclassPreparatoryProduct is broken down into its child classes. The lowest child class in the hierarchy represents a level 4 Uniclass product, e.g. a submersible pump (has code Uniclass code Pr_65_53_24_86). Every aso:UniclassPreparatoryProduct has a Uniclass code through the data property aso:hasUniclassCode. The property aso:hasUniclassCode has its domain as aso:UniclassPreparatoryProduct, and thus, any instance that has this property will be inferred to be of type aso:UniclassPreparatoryProduct.

```
      aso:UniclassPreparatoryProduct rdfs:subClassOf aso:Component ;
        owl:equivalentClass [rdf:type owl:Restriction;
                             owl:onProperty aso:hasSerialNumber;
                             owl:allValuesFrom aso:SerialNumber].
```

LISTING 2
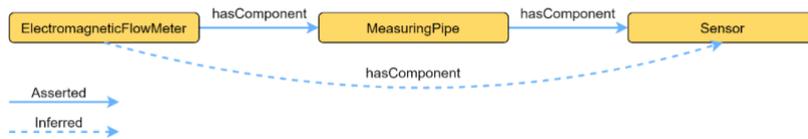

FIG 3: Direct subclasses of aso:Component

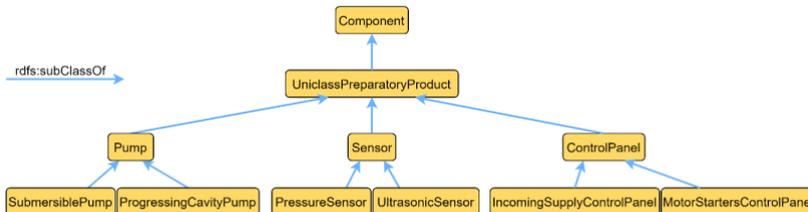*FIG 4: An example of a transitive property*

FIG 5: An example of the component hierarchy

Regarding concept 4, the owl:hasKey built-in property is a tool for specifying the main (key) property of a class. This axiom states that every named instance of a class is identified by a key property or set of properties (Golbreich and Wallace, 2012). Herein, it is necessary to define an aso:Component as much as possible as it is the key class of the ASO. The aso:hasSerialNumber and aso:hasGUID object properties were chosen to be primary key properties, as shown in Listing 3. However, GUIDs are only assigned to assets once they are installed, and so, they will not be of value during procurement.

```
aso:UniclassPreparatoryProduct owl:hasKey (aso:hasGUID),
    (aso:hasSerialNumber).
```

LISTING 3

Regarding concept 5, aso:hasFeatures is a generic property that is found in every UniclassPreparatoryProduct. However, product do not always have a value for this property (not mandatory), if they did, the value is usually a general description that is specific to the product. Therefore, aso:hasFeatures was modelled as a data property of UniclassPreparatoryProduct with a universal restriction (allValuesFrom), in Listing 4.

```
aso:hasFeatures rdf:type owl:DatatypeProperty.

aso:UniclassPreparatoryProduct owl:equivalentClass
    [rdf:type owl:Restriction;
     owl:onProperty aso:hasFeatures;
     owl:allValuesFrom xsd:string].
```

LISTING 4

The PDTs include rich contextual information that was modelled in ASO through restrictions. For instance, in the PDT for an Electromagnetic Flow Meter, the property "analogue output" has the Boolean value of either yes or no. However, if the value is "yes", then there must be values for the number and type of analogue outputs, Fig. 6. Such issue cannot be modelled in OWL without the addition of a rule language. This is one of the reasons that SHACL was introduced in the ASO model, which is discussed in section 5. Nonetheless, an OWL class named aso:AnalogueOutput was created with the two enumerated instances aso:AnalogueOutputAvailable and aso:AnalogueOutputNotAvailable via owl:oneOf, which locally closes the world (Allemang and Hendler, 2011). These axioms which are shown in Listing 5 will be later used with SHACL to implement the IF-THEN rule.

```
aso:AnalogueOutput rdf:type owl:Class;
    owl:equivalentClass [rdf:type owl:Restriction;
                          owl:oneOf (aso:AnalogueOutputAvailable
                                aso: AnalogueOutputNotAvailable)].
```
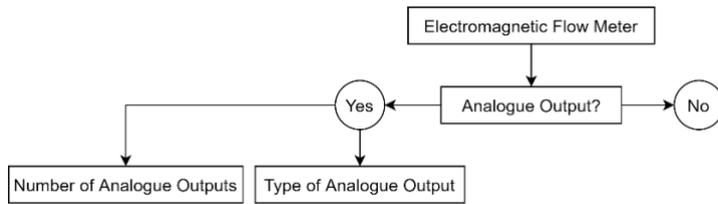
LISTING 5

FIG 6: Example of restrictions

As for concepts 6 and 7, the optimum flow rate of a pump is a good example of a property that requires a QUDT unit and could be linked to OPM. The PDT for a submersible pump includes an optimum flow rate property, which is measured at Best Efficient Point and maximum rotational speed. However, other pumps have different measurement restrictions. For instance, a progressing cavity pump's optimum flow rate is measured at zero discharge pressure, zero suction lift, and fluid temperature of 20 °C. This was modelled in OWL by creating an abstract class called aso:PumpOptimumFlowRate that has restrictions specifying the data type and unit of its instances and asserting the optimum flow rate of each pump type as a subclass of aso:PumpOptimumFlowRate as shown in Listing 6. This modelling paradigm would aid inferring new knowledge from the data rather than act as a data validation tool.

```
aso:PumpOptimumFlowRate rdf:type owl:Class;
  rdfs:subClassOf [rdf:type owl:Restriction;
                    owl:onProperty qudt:unit;
                    owl:hasValue qudt:L-PER-SEC],
                  [rdf:type owl:Restriction;
                    owl:onProperty aso:hasValue;
                    owl:qualifiedCardinality "1"^^xsd:nonNegativeInteger
                    owl:onDataRange xsd:float].

aso:SubmersiblePumpOptimumFlowRate rdfs:subClassOf aso:PumpFlowRate,
    [rdf:type owl:Restriction;
     owl:onClass aso:BestEfficientPoint;
     owl:onProperty aso:measuredAt;
     owl:qualifiedCardinality "1"^^xsd:nonNegativeInteger],
    [rdf:type owl:Restriction;
     owl:onClass aso:UpperLimitRotationalSpeed;
     owl:onProperty aso:measuredAt;
     owl:qualifiedCardinality "1"^^xsd:nonNegativeInteger],
```

LISTING 6

4.4 Reuse of existing terms and ontologies

According to Allemang and Hendler (2011) "OWL should be considered as a living language, growing in the context of the ways it is being used on the web and in commerce". Having decentralised knowledge models with inferring and reasoning capabilities is ought to be of high value in delivering the Industry 4.0 goals (Jaskó et al., 2020). It is possible to have different ontologies modelling the same knowledge or domain, resulting in inconsistencies in data input, which would result in interoperability problems, whether on a single project scale or multiple projects. Ontologies are ought to overcome this through *ontology alignment* (Gómez-Pérez et al., 2004). ontology alignment can be great help when mapping several ontologies or semantic data sets (Cheng et al., 2008). Hence, before designing the ASO, a literature review and search for existing ontologies was performed to identify matching concepts. No existing ontologies for water infrastructure's assets' specification were found. Following Semantic Web best practices and guidelines in linking to existing ontologies and terminologies where possible, the ASO reuses some of the concepts and terminologies from QUDT and dbpedia.org. Linking the ontology to published ontologies is important for achieving Linked Data. For example, the "Gross Weight" property is required for every product in the PDT and linking that with the QUDT ontology makes the distinction that this property

refs to mass and not weight. The QUDT ontology acts a global reference for units, provides unit conversion services, and aids dimensional analysis (Allemang and Hendler, 2011). It is possible to link the common model to the OPM presented in Rasmussen et al. (2019), as mentioned in the previous sub-sections. Yet, it is not clear how beneficial it would be if suppliers provided metadata for property management (e.g. opm:CurrentPropertyState) during the procurement phase. Nonetheless, the ASO is designed with the whole life cycle of information in mind in line with ISO 23386 (British Standards Institution, 2020) and thus, each feature of interest (FOI) is modelled as a class to aid with ontology alignment.

5   DATA VALIDATION

Ontology-Based Data Access (OBDA) can be used for querying relational databases via an ontology using SPARQL (Akinyemi et al., 2020). OBDA implements a virtual data integration approach, as opposed to a warehousing approach (Kharlamov et al., 2017). Therefore, the suppliers' website would be the data sources, which, hence, are not internal databases at the water company. The ontology in this regard represents the mediator, over which, end user queries are performed on the data source. This task is the most complex, as accurate and parametric Information Retrieval (IE) is non-trivial. Product suppliers for the water industry tend to, publish product information in different structures and formats. Maedche et al. (2003) stated that with data integration systems that involve extracting structured data from Natural Language (NL), the more specific the domain is, the less reusable the IE system becomes. In such systems, Information Technology (IT) staff become the 'go to' mediators between the data source and the end user. This is the case for large and data intensive companies in general (Kharlamov et al., 2017). Also, due to the level of detail of the ontology and the randomness of online product information, making such system a generic tool for the water industry may not be feasible.

Therefore, standardising common models for product data exchange seems promising, particularly that end users can be enabled to formulate queries via APIs (Rasmussen et al., 2019). In this regard, data validation becomes the next issue, which was highlighted as an important aspect of product data exchange during the focus groups. Data validation is an important aspect for any enterprise; having a well modelled data structure (schema) is certainly beneficial but cannot be fully exploited without a mechanism to validate the quality of the data (Labra Gayo et al., 2018). Generally, modellers should choose between open and closed "world assumption" (Böhms et al., 2009). Although a DL ontology can be considered as a data validation tool in itself – ontology-based validation (Labra Gayo et al., 2018) through logical axioms – the Open World Assumption (OWA) makes it unrealistic for the strict data validation needed in AEC industry. The mantra for the Semantic Web is that "Anyone can say Anything about Any topic" (Allemang and Hendler, 2011), which is not very useful in situations where data quality assurance is important. For example, the motor in a submersible pump must have an insulation class rating of either Y, A, E, B, F, H, N, or R. Thus, it should not be possible to add any other ratings since these ratings have been define by BS EN 60085 for the UK industry (British Standards Institution, 2008). Hence, it is necessary to declare OWL individuals within the same files to be different from one another using owl:differentFrom (Barbau et al., 2012). However, for the model in this research with a relatively large number of instances, other validation methods are necessary. Additionally, using OWL restrictions such as owl:allValuesFrom and owl:someValuesFrom albeit adds semantic richness to the data, it aids inferencing rather than validation. There is a need for other validation tools to aid product requirements checking and comparison.

RDF data validation can be done using SPARQL. However, SPARQL can be verbose and exhaustive if it was to be used for checking large and complex data structures. SPARQL Inferencing Notation (SPIN) was proposed as a mechanism to attach SPARQL-based rules and constraints to classes. Zhang et al. (2018) proposed extending SPARQL queries by using SPIN functions, for instance, to query ifcOWL data, as they claimed that industry practitioners working with IFC instance building models face data quality issues when retrieving domain-specific information. However, SPIN is hard to read and maintain by nonexperts (Labra Gayo et al., 2018). In this regard, two languages that were investigated in this research were Shape Expressions (ShEx) and Shapes Constraint Language (SHACL). ShEx was developed by the ShEx Community Group (Prud'hommeaux at al., 2019) and SHACL was developed by the RDF Data Shapes Working Group (RDF Data Shapes Working Group Charter, 2017). Labra Gayo et al. (2018) Provided a comparison between ShEx and SHACL, and from their discussion it was realised that due to SHACL's expressivity and verboseness, it introduces more delicacy when modelling restrictions in this research. Ontologies are knowledge representation tools (El-Diraby and Osman, 2011), and so, the ASO represents the knowledge that describes a product during procurement at the O&M phase. SHACL on

the other hand, is ought to introduce restrictions on this knowledge set by the end user, based on which, information retrieval can be applied. For example, if an end user wanted to procure a submersible pump with no less than 100 units for power rating. Otherwise, the cost of the pump should be less than 5000 units. Listing 7 shows such a constraint.

```
aso:SubmersiblePumpShape rdf:type sh:NodeShape;
   sh:targetClass aso:SubmersiblePump;
   sh:or ([sh:property [
        sh:path ([sh:oneOrMorePath aso:hasUpperLimitContinuousPowerRating] aso:hasValue);
        sh:minExclusive 100]]


      [sh:property [
        sh:path ([sh:oneOrMorePath aso:hasCost] aso:hasValue);
        sh:maxExclusive 5000 ]]).
```

LISTING 7

IF-THEN restrictions can be modelled in SHACL using NOT and OR operators. For example, IF (x) THEN (y) can be modelled as NOT (x) OR (y). Likewise, IF AND ONLY IF (x) THEN (y) can be modelled as ((NOT (x) OR (y)) AND ((NOT (y) OR (x)). Using SHACL, the electromagnetic flow meter example from sub-section 4.3 can be modelled in Turtle syntax as shown in Listing 8.

```
aso:ElectromagneticFlowMeterShape rdf:type sh:NodeShape;
    sh:targetClass aso:ElectromagneticFlowMeter;
    sh:property [
        sh:path aso:hasAnalogueOutput;
        sh:in (aso:AnalogueOutputAvailable aso:AnalogueOutputNotAvailable);
        sh:maxCount 1;
        sh:minCount 1;];

    sh:or (
      [sh:not [
        sh:path aso:hasAnalogueOutput;
        sh:hasValue aso:AnalogueOutputAvailable;]]
      [sh:and ([sh:path aso:hasTypeOfAnalogueOutputs;
               sh:datatype xsd:string;
               sh:maxCount 1; sh:minCount 1;]
             [sh:path aso:hasNumberOfAnalogueOutputs;
               sh:datatype xsd:integer;
               sh:maxCount 1; sh:minCount 1;])]);

    sh:or (
      [sh:path aso:hasAnalogueOutput;
       sh:hasValue aso:AnalogueOutputAvailable;]
      [sh:not  [sh:and (
        [sh:path aso:hasTypeOfAnalogueOutputs;
         sh:datatype xsd:string;
         sh:maxCount 1; sh:minCount 1;]
        [sh:path aso:hasNumberOfAnalogueOutputs;
         sh:datatype xsd:integer;
         sh:maxCount 1; sh:minCount 1;])]]).
```

LISTING 8

It is important to consider that SHACL should be applied after inference has taken place, as SHACL validators do not have full inference capabilities. Also, declarative documentation should be published to serve as a contract between data producers and modellers – users of the model (Labra Gayo et al., 2018).

SHACL rules were added to the ontology through a SHACL plug-in to Protégé. However, SHACL code must be written manually. Both SHACL core and SHACL-SPARQL were utilised in this model. Using SHACL-SPARQL requires an underlying SPARQL processor. It is important to note that SHACL implementation differs over the project's lifecycle, due to the differing constraints required by different stakeholders. Thus, there should be phase-specific implementations of SHACL constraints.

6   APPLICATION OF ASO

The ISO 23386 (British Standards Institution, 2020) recommends distributed data dictionaries for BIM, and so, "different groups, possibly in different countries will create or have created separate data dictionaries, specialised for their needs, based on the legislation and culture". Hence, an ontological approach was proposed due to the relative ease of mapping properties to classes using description logic. The Common Data Environment (CDE) presented in PD 19650-0:2019 (British Standards Institution, 2019) does not specify the kind of information required in a CDE in detail. In the case of the water industry for the UK, there is some consensus on the required asset data to be exchanged during the O&M phase manifested in the PDTs, and hence, the common model presented here is ought to collect and validate asset data requirements as part of the Project Information Model (PIM) and Asset Information Model (AIM). Therefore, from a standards point of view, the value of this common ontology model is to deliver a more standardised and systematic asset information during product procurement which can yield cost and time savings to the overall delivery of the project.

Extendibility should be taken into consideration when the ontology is being designed and developed. The abstract class "aso:Component" is the key class of the ASO, and therefore, it should be the starting point for extensions. For instance, an ASO for the water domain and an ASO for the oil and gas or railway domains would both include the component class, assuming that they share the same abstraction principles. The aspiration is to have the ASO as part of a semantic common model library that contains product data knowledge representation populated by suppliers, Fig. 7. SHACL APIs would allow end users (i.e. designers) to customise the restrictions and query the knowledgebase. Additionally, results from the focus group sessions suggested that highly detailed common models should be phase-specific, due to the differing requirements from one project phase to another, and therefore, this research focuses on the O&M phase. However, such phase-specific common models must fit within the context of the Whole Life Cycle (WLC) of information flow (Dawood and Vukovic, 2015).
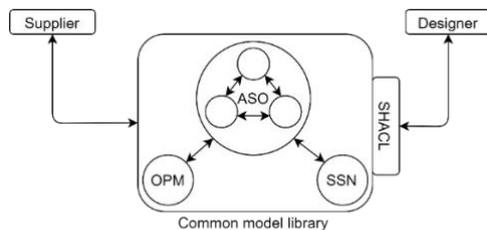

FIG 7: Common model library for product data

7   DISCUSSION

7.1   Requirements capture

It was realised during the focus group sessions that one of the current inefficiencies of data exchange within water infrastructure in the UK is the differing asset classification methods and requirements among water companies. This leads to uncertainty in the supply chain on what asset data are required and how to structure them (Wu et al., 2019). Classification differences are not only observed among companies, but among the different stages of a project within the same company. Therefore, the process of asset procurement becomes a lengthy negotiation process involving several parties of the supply chain to ensure a common understanding. In this regard, the PDTs were a national effort to standardise data requirements and structure for the UK water industry. Although the PDTs are a step in the right direction from a data dictionary point of view, they do not provide the semantic relationships offered by a common model, neither do they provide mechanisms of data validation. PDTs rely on Uniclass for

asset hierarchical classification, which has logical contradictions as discussed in sub-section 7.2. During the focus groups sessions, when participants were shown some of these contradictions, they agreed on that Uniclass needs major revisions. Along similar lines, there are no current standards that go beyond asset naming convention to name individual properties. For instance, level four of the Uniclass structure, which is the most detailed level, only classifies submersible pumps as Pr_65_53_24_86, without any properties; hence the need for data dictionaries such as PDTs to be standardised. Additionally, the internal classification system used by water companies in the UK can be ambiguous. For example, the term "asset" is sometimes used to refer to a process, and the term "sub-asset" sometimes refers to a physical asset. On other hand, some water companies' asset hierarchy identifies site types rather than physical equipment. Such representations have been carried over to the Uniclass system. Within this context, respondents encouraged further standardisation as well as standardising certain important asset functional properties such as flow rate, and energy consumption. Functional properties (not to be confused with OWL functional properties) are particular to the asset's functions within the water company. On the other hand, it was also realised that the PDTs offer more detailed properties than most water companies in the UK require from an operations and maintenance (O&M) perspective. This suggests that despite the PDTs initiative, the O&M sector for most water companies in the UK is not mature compared to other phases of the project lifecycle. However, Participants agreed on that the semantic meaning brought by an OWL common model would be welcomed. For example, having the ability to track a specific "seal" component to a specific "submersible pump" is of value.

Another point that was raised during the sessions was that data input is poor and there is no mechanism to evaluate it. The importance of data validation was considered a major issue currently facing the industry. A first step towards validation would be setting thresholds on the data. Participants agreed that this would aid semantic understanding of the requirements. Therefore, founding the asset procurement negotiation process through a standardised common model that can set constraints on the data is ought to be beneficial. The gathered data suggested that SHACL should be used in conjunction with the underlaying ontology framework. Generally, participants agreed on the need to develop a common process for the exchange of product data for the water industry. This provides confirmatory evidence common models form a step forward in the right direction. Nonetheless, having common models as sources of truth requires extensive testing and validation of the models themselves.

Generally, participants were keen on mapping classifications to the common model, which would aid semantic meaning. Participants agreed on that classifications need to be converted into a list of properties, and then, establish connections between them and the properties in the common model. Description logic semantics would aid this objective from a Linked Data perspective. For example, mapping the ASO to the OPM (Rasmussen et al., 2019) would provide some semantic link to ISO 23386 (British Standards Institution, 2020). OPM would need to be updated to fully accommodate ISO 23386, however. Additionally, participants agreed on the importance of data dictionaries as a pre-common model step, as some standards lack contextual information. An ISO standard property could have different lists of subproperties that belong to different domains and certain properties may have different names in different standards. In this regard, having several standards competing for managing construction data adversely affects interoperability (Ozturk 2020; Shen et al. 2010). Therefore, data dictionaries are important in mapping between classifications.

Respondents also shared the opinion that the common model must be created from existing standards, definitions, objects, and properties, as the general aspiration of the UK water industry is to abstract the business entities and requirements to a modular water digital twin. Thus, every water company would a have context-specific digital twin that could be mapped through standardised classifications to a national digital twin. Based on this, one can assume that the relevant material already exists, and yet, the problem lies in making semantic connections between what is available. Generally, the benefits of adding semantic meaning could be summarised as follows:

- Mapping standards to properties – Properties could be shared by more than one industry
- Mapping standards to standards – Different naming conventions are sometimes used for the same property in different standards
- Mapping standards to dictionaries – Reference data libraries could be ambiguous

Fig. 8 summarises the resultant method that emerged from the focus group sessions for a data exchange scenario, which would be applied for every possible data exchange instance using SWT. This results in a model library that would include common models for product procurement at different phases of the project, and hence, different data exchange scenarios. Users would choose a suitable common model based on the business requirements for

product procurement. Therefore, it is important to consider the whole life cycle of information when creating any particular property, which is where ISO 23386 becomes relevant (British Standards Institution, 2020).
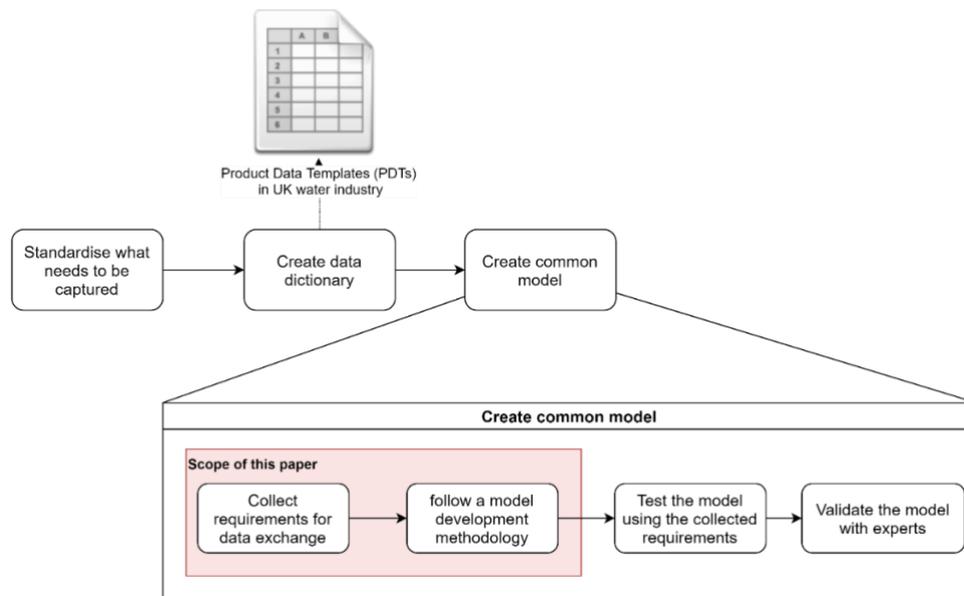

FIG 8: Scenario-specific data exchange framework

7.2   Common model

A main contribution to knowledge by this ontology is portrayed in the size of it. Most published ontologies for product data exchange are general top-level ontologies. These ontologies are important as they establish the semantic foundations for the alignment of application ontologies. Nevertheless, extended application ontologies that represent real life scenarios in this domain were not observed. An application ontology is a detailed extension of a domain ontology resulting in a specific subdomain. Application ontologies connect a philosophical phenomenon with the digital paradigm of AI tools (El-Diraby and Osman, 2011). To this end, one of the modelling issues that arose when developing ASO was an "illogical loop" caused by adopting the Uniclass system. In this common model example, a sensor – which is a Uniclass product – has two subclasses; ultrasonic sensor and pressure sensor. However, an electromagnetic flow meter was also modelled as a Uniclass product and has a component that is a sensor. This would result in the electromagnetic flow meter having a component that is a Uniclass sensor, which is not the case, as this type of sensors (component of an electromagnetic flow meter) is not specified by Uniclass, and hence, does not have a PDT. In addition, a control panel was initially modelled as a Uniclass product, having subclasses such as incoming supply control panel, instrumentation automation and control panel, and motor starters control panel. However, Uniclass has classified another product as "control panels" with the code Pr_75_50_18_17, meaning that it should also fall under the superclass control panel. This causes an illogical semantic representation, which indicates that the Uniclass classification system was not designed with modelling in mind, as it classifies products by general function (functional modality) on some occasions and by domain of use (sector modality) on others. Uniclass was made based on current industry practices which may have magnified some of the issues when modelling products in the water domain. Additionally, some inconsistencies can occur when modelling properties from a data dictionary. For example, "Type" property that is specified by bsDD may have a different meaning to rdf:type axioms.

Due to the semantics OWL provides and the ability to map ontologies to each other and to standards, it is believed that OWL common models can bring benefits to the issues discussed in this section, and hence, enhance the process of data management and exchange. Top-level ontologies are cross domain and holistic view of the things (El-Diraby and Osman, 2011), and hence, they provide some guidance on how to develop detailed ontologies for the industry. Top-level ontologies can be used as the alignment tools that mediate between the decentralised knowledge systems. In this regard, the ASO can be regarded as a representation ontology according to Martinez-Cruz et al. (2012) and an application ontology according to El-Diraby and Osman (2011). However, it is not

feasible for one model to represent diverse concepts, which is why it is suggested to have a network of common models for the water infrastructure. The aspiration here is have an ontological network (model libraries) in a similar manner to the OBO (Open Biological and Biomedical Ontology) Foundry (The OBO Foundry, 2021). A semantic common model library for the water industry's infrastructure assets would not only allow product information extraction, but also requirements comparison and product matching. It is important to highlight not just the immediate technical benefits of such knowledgebases for data requirements capture, but also the expected benefits throughout the product lifecycle. Supplying the client/operator with better quality, structured, computer and human readable, and query-able data would result in cost savings over the assets' life cycle. Moreover, it is important to have a reference system for naming and classifying different assets. The Uniclass naming convention was initially established to do that and reduce ambiguity. However, as discussed in sub-section 7.1, Uniclass was not created for semantic modelling, and therefore, if semantic common models were to be adopted, Uniclass should be restructured.

As mentioned in section 1, knowledge graphs are potentially the future of BIM as they provide schema independent models (Bradley et al., 2016), which may be true for RDF data. However, as seen with the development of the ASO, rich and complex DL models require restrictions and even rules, which eventually results in a schema-dependent model. Implementing the ontology in OWL DL is useful for reasoning and inferencing. However, some triple stores may restrict the ontology to other OWL 2 subsets (RL, QL, or EL) or even RDFS-Plus, resulting in reduced expressivity. Also, since research on SWT in AEC have only started accumulating in the past decade (Zhong et al., 2019) , it was noticed that the limitations of SWT in AEC are infrequently discussed. For instance, case studies evaluating the performance of complex ontological models in AEC are scarce.

The Open World Assumption (OWA) of the Semantic Web may not be the perfect tool for modelling information requirements. For instance, in the electromagnetic flow meter example in Fig. 6, despite that the "AnalogueOutput" class was enumerated with the only two possible instances (AnalogueOutputAvailable and AnalogueOutputNotAvailable), the model would still allow other instances to be added to the class, unless one explicitly asserts that the two instances are owl:differentFrom every other possible instance. This is due to the nonunique naming assumption of the OWL language (Allemang and Hendler, 2011). Also, one may want to assert that a certain class has only a certain number of child classes, which is not possible in OWL, primarily because it was designed with the OWA – to be as realistic as possible in modelling real phenomenon from a logical and philosophical viewpoint. For instance, if a model consisted of only one type of pumps, say a submersible pump, then it is not possible to assert that class "Pump" only has one subclass, that is a "SubmersiblePump". This is because the model will always 'assume' that new information can be added anytime. Therefore, using SHACL albeit useful in such situations, it changes the characteristics initially set for the OWL language.

Some of the assets in ASO are shared by other domains such as the oil and gas domain, and hence, they may have different criteria to describe their assets than the water industry. Nonetheless, due to the "linking across domains" potential of the Semantic Web (Pauwels et al., 2017), other domains can extend this ontology and populate it to their needs. This paper only presented the main concepts of the proposed ontology (ASO) and future work will focus on modularising it; converting the many parts of this ontology into smaller, specific, properly managed ontologies. For instance, in Listing 9, there could be a dedicated ontology for the "enclosure rating" class. This is because with detailed ontologies, maintaining modularity becomes an issue, and consequently, importing parts of the ASO would result in non-related functionalities (Wagner and Rüppel, 2019). Another issue that arose in this research is the definition of the minimum required properties for certain assets. For example, it is still not clear which properties should have existential restrictions and which should have universal restrictions. This could form the topic of future research.

```
aso:SubmersiblePump rdfs:subClassOf
  [rdf:type owl:Restriction;
   owl:onProperty aso:hasEnclosureRating;
   owl:qualifiedCardinality "1"^^xsd:nonNegativeInteger;
   owl:onClass aso:EnclosureRating].
```

LISTING 9

8 CONCLUSION

Current product procurement practices in the UK water industry involve static PDTs with no means of data querying or validation. To this end, this research has conducted focus group sessions with industry experts in the UK to identify problems and requirements. Then, a semantic common model was proposed to address the problems based on the identified requirements, as part of an ongoing research.

Results from these sessions suggested that although there are efforts to standardise asset hierarchies, classifications, and data exchange for the UK water industry, current practices for exchange and management of product data are affected by the lack of interoperability. The reasons for this lack of interoperability include bad data structures, lack of software solutions, lack of participation from suppliers, and the lack of unified classifications and processes. The main conclusion that emerged from the focus group sessions is that *information exchange in the water industry in the UK is impeded due to the lack of structured and semantic data*. From these sessions, the collected requirements were used to create an ontological common model for product information capture and validation named the ASO. The current process of acquiring asset information is done manually via PDTs, which are becoming normative in the United Kingdom, at least within the water industry. Therefore, creating an ASO from the PDTs would include axioms that represent consensus knowledge in this domain. Participants generally agreed on the fact that there needs to be semantic mapping from a) standards to properties, b) standards to standards, and c) standards to dictionaries.

OWL ontologies seem like a promising technology in achieving interoperability. However, it was generally noticed in the literature that limitations of the SWT as modelling tools for AEC are seldom discussed. According to Pauwels et al. (2015), research is needed on the technical issues that are slowing the implementation and deployment of semantic web technologies in the construction industry. Data validation, which emerged as a requirement from the focus group sessions, requires closing the world, at least in certain scenarios to be set by the end user. Within the context of this study, an OWL common model for product data was not enough for the model to be complete. Hence, combing OWL's semantic expressivity with SHACL rules and constraints is deemed useful in achieving the objectives of this research. The common model framework presented in this paper suggested that whilst the data schema (OWL) would be the same for the industry, SHACL restrictions can be customisable modules by each company for suitability of procurement and operation. This framework can be regarded as the first iteration of a prototype, that is ought to be validated by domain experts at a later stage of this research. It is worth noting that the solution to the problem introduced in this paper is non-trivial, as it would require efforts from ontology engineers, data engineers, and domain experts.

9 LIMITATIONS AND FUTURE WORK

A limitation to this study was the relatively small number of experts for the requirements capture process. Albeit the target sample is relatively small, future work should consider a wider sample of participants for deductive reasoning of the results. Also, Future work will focus on modularising the ontology and conducting Information Retrieval (IR) based on SHACL constraints to evaluate the proposed system in Fig. 7.

REFERENCES

Abdul-Ghafour, S., Ghodous, P., Shariat, B. and Perna, E. (2007). A Common Design-Features Ontology for Product Data Semantics Interoperability. In: *IEEE/WIC/ACM International Conference on Web Intelligence*. IEEE, pp.443-446.

Abdul-Ghafour, S., Ghodous, P., Shariat, B., Perna, E. and Khosrowshahi, F. (2014). Semantic interoperability of knowledge in feature-based CAD models. *Computer-Aided Design*, 56, pp.45-57.

Akinyemi, A., Sun, M. and doing, A. (2020). Data integration for offshore decommissioning waste management. *Automation in Construction*, 109, p.103010.

Allemang, D. and Hendler, J. (2011). *Semantic web for the working ontologist*. 2nd ed. Morgan Kaufmann.

Beetz, J. and Vries, D. (2009). Building product catalogues on the Semantic Web. *Managing IT in Construction/Managing Construction for Tomorrow*, pp.239-244.

Beetz, J., van Leeuwen, J. and de Vries, B. (2009). IfcOWL: A case of transforming EXPRESS schemas into ontologies. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 23(1), pp.89-101.

Berners-Lee, T., Bray, T., Connolly, D., Cotton, P., Fielding, R., Jeckle, M., Lilley, C., Mendelsohn, N., Orchard, D., Walsh, N. and Williams, S. (2004). *Architecture Of The World Wide Web, Volume One*. [online] W3.org. Available at: <https://www.w3.org/TR/webarch/#id-resources> [Accessed 5 May 2020].

Bilal, M., Oyedele, L., Munir, K., Ajayi, S., Akinade, O., Owolabi, H. and Alaka, H. (2017). The application of web of data technologies in building materials information modelling for construction waste analytics. *Sustainable Materials and Technologies*, 11, pp.28-37.

BIM4Water Standard Libraries Group. (2019). The Development Procedure For Water Industry Product Data Templates. BIM4WATER.

Borst, P. (1997). *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. Enschede: Centre for Telematics and Information Technology.

Böhms, M., Bonsma, P., Bourdeau, M., Kazi, A. S. (2009). Semantic product modelling and configuration: challenges and opportunities, ITcon, 14, Special issue Next Generation Construction IT: Technology Foresight, Future Studies, Roadmapping, and Scenario Planning, pg. 507-525.

Bradley, A., Li, H., Lark, R. and Dunn, S. (2016). BIM for infrastructure: An overall review and constructor perspective. *Automation in Construction*, 71, pp.139-152.

British Standards Institution. (2008). *BS EN IEC 60085:2007: Electrical insulation — Thermal evaluation and designation.*

British Standards Institution. (2016). *BS EN ISO 12006-3:2016: Building construction — Organization of information about construction works — Part 3: Framework for object-oriented information (ISO 12006-3:2007).*

British Standards Institution. (2018). *BS EN ISO 19650-1:2018: Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) – Information management using building information modelling.*

British Standards Institution. (2019). *PD 19650-0:2019: Transition guidance to BS EN ISO 19650.*

British Standards Institution. (2020). *BS EN ISO 23386:2020: Building information modelling and other digital processes used in construction – Methodology to describe, author and maintain properties in interconnected data dictionaries.*

British Standards Institution. (2020). *BS EN ISO 23387:2020: Building information modelling (BIM) — Data templates for construction objects used in the life cycle of built assets — Concepts and principles.*

Britishwater.co.uk. (n.d). *British Water.* [online] Available at: <https://www.britishwater.co.uk/bim4water-standards-library.aspx> [Accessed 1 July 2021].

Cavka, H., Staub-French, S. and Poirier, E. (2017). Developing owner information requirements for BIM-enabled project delivery and asset management. *Automation in Construction*, 83, pp.169-183.

Chaparala, R., Hartman, N. and Springer, J. (2013). Examining CAD Interoperability through the Use of Ontologies. *Computer-Aided Design and Applications*, 10(1), pp.83-96.

Cheng, J., Lau, G., Pan, J. and Law, K. (2008). Domain-Specific Ontology Mapping by Corpus-Based Semantic Similarity. In: *NSF CMMI Engineering Research and Innovation Conference*.

Cheng, J., Lu, Q. and Deng, Y. (2016). Analytical review and evaluation of civil information modeling. *Automation in Construction*, 67, pp.31-47.

Cibse.org. n.d. *CIBSE - Building Information Modelling*. [online] Available at: <https://www.cibse.org/Knowledge/BIM-Building-Information-Modelling> [Accessed 4 July 2021].

Corrado, V., Ballarini, I., Madrazo, L. and Nemirovskij, G. (2015). Data structuring for the ontological modelling of urban energy systems: The experience of the SEMANCO project. *Sustainable Cities and Society*, 14, pp.223-235.

Costa, G. and Madrazo, L. (2015), Connecting building component catalogues with BIM models using semantic technologies: An application for precast concrete components, *Automation in Construction*, Elsevier B.V., Vol. 57, pp. 239–248.

Cyganiak, R., Wood, D. and Lanthaler, M. (2014). *RDF 1.1 Concepts And Abstract Syntax*. [online] W3.org. Available at: <https://www.w3.org/TR/rdf11-concepts/> [Accessed 13 August 2020].

Dawood, N. and Vukovic, V. (2015). WHOLE LIFECYCLE INFORMATION FLOW UNDERPINNED BY BIM: TECHNOLOGY, PROCESS, POLICY AND PEOPLE. In: *2nd International Conference on Civil and Building Engineering Informatics*. Tokyo.

Designingbuildings.co.uk. (2020). Product Data Template. [online] Available at: <https://www.designingbuildings.co.uk/wiki/Product_Data_Template> [Accessed 29 June 2021].

Doan, A., Halevy, A. and Ives, Z. (2012). *Principles of Data Integration*. London: Morgan Kaufmann.

El-Diraby, T. and Osman, H. (2011). A domain ontology for construction concepts in urban infrastructure products. *Automation in Construction*, 20(8), pp.1120-1132.

Elsawah, S., Pierce, S., Hamilton, S., van Delden, H., Haase, D., Elmahdi, A. and Jakeman, A. (2017). An overview of the system dynamics process for integrated modelling of socio-ecological systems: Lessons on good modelling practice from five case studies. *Environmental Modelling & Software*, 93, pp.127-145.

Gao, G., Liu, Y.S., Wang, M., Gu, M. and Yong, J.H. (2015), A query expansion method for retrieving online BIM resources based on Industry Foundation Classes, *Automation in Construction*, Elsevier B.V., Vol. 56, pp. 14–25.

Gao, G., Liu, Y.S., Lin, P., Wang, M., Gu, M. and Yong, J.H. (2017), BIMTag: Concept-based automatic semantic annotation of online BIM product resources, *Advanced Engineering Informatics*, Elsevier Ltd, Vol. 31, pp. 48–61.

Golbreich, C. and Wallace, E. (2012). *OWL 2 Web Ontology Language New Features And Rationale (Second Edition)*. [online] W3.org. Available at: <https://www.w3.org/TR/owl2-new-features/> [Accessed 11 August 2020].

Gómez-Pérez, A., Fernández-López, M. and Corcho, O. (2004). *Ontological Engineering: with examples from the areas of knowledge management, e-commerce, and the semantic web*. London: Springer-Verlag.

Gray, D. (2018). *Doing Research In The Real World*. 4th ed. Sage.

Gruber, T. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), pp.199-220.

Grüninger, M. and Fox, M. (1995). Methodology for the Design and Evaluation of Ontologies. In: *IJCAI-95*.

Hippolyte, J., Rezgui, Y., Li, H., Jayan, B. and Howell, S. (2018). Ontology-driven development of web services to support district energy applications. *Automation in Construction*, 86, pp.210-225.

Howell, S., Rezgui, Y., Hippolyte, J., Jayan, B. and Li, H. (2017). Towards the next generation of smart grids: Semantic and holonic multi-agent management of distributed energy resources. *Renewable and Sustainable Energy Reviews*, 77, pp.193-214.

Jaskó, S., Skrop, A., Holczinger, T., Chován, T. and Abonyi, J. (2020). Development of manufacturing execution systems in accordance with Industry 4.0 requirements: A review of standard- and ontology-based methodologies and tools. *Computers in Industry*, 123.

Kharlamov, E., Bilidas, D., Hovland, D., Jimenez-Ruiz, E., Lanti, D., Lie, H., Rezk, M., Skjjveland, M., Soylu, A., Xiao, G., Zheleznyakov, D., Giese, M., Ioannidis, Y., Kotidis, Y., Koubarakis, M. and Waaler, A. (2017). Ontology Based Data Access in Statoil. *Journal of Web Semantics*, 44, pp.3-36.

Kim, K., Kim, H., Kim, W., Kim, C., Kim, J. and Yu, J. (2018). Integration of ifc objects and facility management work information using Semantic Web. *Automation in Construction*, 87, pp.173-187.

Kuster, C., Hippolyte, J. and Rezgui, Y. (2020). The UDSA ontology: An ontology to support real time urban sustainability assessment. *Advances in Engineering Software*, 140.

Labra Gayo, J., Prud'hommeaux, E., Boneva, I. and Kontokostas, D. (2018). *Validating RDF Data*. Morgan and Claypool.

Li, Y., García-Castro, R., Mihindukulasooriya, N., O'Donnell, J. and Vega-Sánchez, S. (2019). Enhancing energy management at district and building levels via an EM-KPI ontology. *Automation in Construction*, 99, pp.152-167.

Li, Z. and Ramani, K., 2007. Ontology-based design information extraction and retrieval. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 21(2), pp.137-154.

Maedche, A., Neumann, G. and Staab, S. (2003). Bootstrapping an Ontology-based Information Extraction System. In: P. Szczepaniak, J. Segovia, J. Kacprzyk and L. Zadeh, ed., *Intelligent Exploration of the Web*. Physica-Vertag, pp.345-359.

Martinez-Cruz, C., Blanco, I. and Vila, M. (2012). Ontologies versus relational databases: are they so different? A comparison. *Artificial Intelligence Review*, 38(4), pp.271-290.

Naoum, S. (2014). *Dissertation research & writing for construction students*. 2nd ed. London: Routledge.

Obofoundry.org. 2021. *The OBO Foundry*. [online] Available at: <http://www.obofoundry.org/> [Accessed 28 January 2021].

Ozturk, G. (2020). Interoperability in building information modeling for AECO/FM industry. *Automation in Construction*, 113.

Pauwels, P. and Terkaj, W. (2016). EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. *Automation in Construction*, 63, pp.100-133.

Pauwels, P., Krijnen, T., Terkaj, W. and Beetz, J. (2017). Enhancing the ifcOWL ontology with an alternative representation for geometric data. *Automation in Construction*, 80, pp.77-94.

Pauwels, P., Törmä, S., Beetz, J., Weise, M. and Liebich, T. (2015). Linked Data in Architecture and Construction. *Automation in Construction*, 57, pp.175-177.

Pauwels, P., Zhang, S. and Lee, Y.C. (2017), Semantic web technologies in AEC industry: A literature overview, *Automation in Construction*, Elsevier B.V., Vol. 73, pp. 145–165.

Powell, R., Single, H. and Lloyd, K. (1996). Focus Groups in Mental Health Research: Enhancing the Validity of User and Provider Questionnaires. *International Journal of Social Psychiatry*, 42(3), pp.193-206.

Prud'hommeaux, E., Boneva, I., Labra Gayo, J. and Kellogg, G. (2019). *Shape Expressions Language 2.1*. [online] Shex.io. Available at: <https://shex.io/shex-semantics/> [Accessed 14 August 2020].

POWDER Working Group. (2010). *Protocol For Web Description Resources (POWDER): POWDER-S Vocabulary (WDRS)*. [online] W3.org. Available at: <https://www.w3.org/2007/05/powder-s#Document> [Accessed 20 July 2020].

Protege.stanford.edu. (2020). *Protégé*. [online] Available at: <https://protege.stanford.edu/> [Accessed 11 August 2020].

Qudt.org. (2021). *QUDT*. [online] Available at: <http://www.qudt.org/> [Accessed 12 January 2021].

Radulovic, F., Poveda-Villalón, M., Vila-Suero, D., Rodríguez-Doncel, V., García-Castro, R. and Gómez-Pérez, A. (2015). Guidelines for Linked Data generation and publication: An example in building energy consumption. *Automation in Construction*, 57, pp.178-187.

Rasmussen, M., Lefrançois, M., Pauwels, P., Hviid, C. and Karlshøj, J. (2019). Managing interrelated project information in AEC Knowledge Graphs. *Automation in Construction*, 108.

RIBA. (2020). Plan Of Work Overview. [online] London: RIBA. Available at: <https://www.architecture.com/knowledge-and-resources/resources-landing-page/riba-plan-of-work#available-resources> [Accessed 23 January 2020].

Sauermann, L. and Cyganiak, R. (2008). *Cool Uris For The Semantic Web*. [online] W3.org. Available at: <https://www.w3.org/TR/cooluris/> [Accessed 5 May 2020].

Seal, D., Bogart, L. and Ehrhardt, A. (1998). Small group dynamics: The utility of focus group discussions as a research method. *Group Dynamics: Theory, Research, and Practice*, 2(4), pp.253-266.

Standards.buildingsmart.org. (2020). IFC4 Documentation. [online] Available at: <https://standards.buildingsmart.org/IFC/RELEASE/IFC4/ADD2_TC1/HTML/> [Accessed 22 February 2020].

Suárez-Figueroa, M., Gómez-Pérez, A., Motta, E. and Gangemi, A. (2012). *Ontology Engineering In A Networked World*. London: Springer.

Tchouanguem Djuedja, J., Abanda, F., Kamsu-Foguem, B., Pauwels, P., Magniont, C. and Karray, M. (2021). An integrated Linked Building Data system: AEC industry case. *Advances in Engineering Software*, 152.

Uschold, M. and King, M. (1995). Towards A Methodology For Building Ontologies. AIAI-TR-183. University of Edinburgh.

Wagner, A. and Rüppel, U. (2019). BPO: The Building Product Ontology for Assembled Products. In: *Linked Data in Architecture and Construction*. Lisbon.

Wu, S., Shen, Q., Deng, Y. and Cheng, J. (2019). Natural-language-based intelligent retrieval engine for BIM object database. *Computers in Industry*, 108, pp.73-88.

Wang, M., Deng, Y., Won, J. and Cheng, J. (2019). An integrated underground utility management and decision support based on BIM and GIS. *Automation in Construction*, 107.

W3.org. (2017). *RDF Data Shapes Working Group Charter*. [online] Available at: <https://www.w3.org/2014/data-shapes/charter> [Accessed 2 August 2020].

Yuce, B. and Rezgui, Y. (2017). An ANN-GA Semantic Rule-Based System to Reduce the Gap Between Predicted and Actual Energy Consumption in Buildings. *IEEE Transactions on Automation Science and Engineering*, 14(3), pp.1351-1363.

Zhang, C., Beetz, J. and Vries, D. (2018). "BimSPARQL: Domain-specific functional SPARQL extensions for querying RDF building data". *Semantic Web*, 9(6), pp.829-855.