Alani Y., Dawood N., Rodriguez S. and Dawood H. (2020). "Whole life cycle construction information flow using semantic web technologies: a case for infrastructure projects" In: *Proc. 37th CIB W78 Information Technology for Construction Conference* (CIB W78), São Paulo, Brazil, pp. 141-155. DOI: http://dx.doi.org/10.46421/2706-6568.37.2020.paper011

WHOLE LIFE CYCLE CONSTRUCTION INFORMATION FLOW USING SEMANTIC WEB TECHNOLOGIES: A CASE FOR INFRASTRUCTURE PROJECTS

Yasir Alani[1], Nashwan Dawood[2], Sergio Rodriguez[3], and Huda Dawood[4]

Abstract: The construction industry is considered to be lagging behind other industries in terms of the technological advancement. One of the main factors is the lack of integration of incoherent and heterogeneous data on a project level. Whilst the adoption of Building Information Modelling (BIM) technologies and processes was aimed to solve integration issues. The interoperability is still a problem to solve, as most information and data fields show inconsistencies in a number of cases. One of the aspects of the problem is that IFC EXPRESS schema is only machine readable, requires extensive mappings, and usually does not support infrastructure domains other than buildings.

This research explores the possibility of utilising Semantic Web Technologies (SWT) to help achieving some of the desired goals of data interoperability and Whole Life Cycle (WLC) information flow. SWT support the creation of comprehensive, layered, shared, human and machines readable, and extendable knowledge repositories dubbed ontologies. The Resource Description Framework (RDF) which forms the core of SWT provides a rather elegant way of modelling datasets, that is, assigning an Internationalised Resource Identifier (IRI) to each class, instance, and property. SWT are ought to provide better information retrieval and inferencing than current systems used in the industry.

The main objective of this paper is to present a framework to demonstrate how SWT can underpin WLC information flow in water infrastructure projects case study.

Keywords: BIM, Ontology, RDF, SWT, WLC.

1 Introduction

1.1 Building Information Modelling (BIM)

Effective collaboration and systems integration are of significant importance in the Architecture, Engineering, and Construction (AEC) industry. The fragmented characteristics of this industry makes it very difficult to fully exploit the benefits of Information and Communication Technology (ICT), in comparison to other industries (Li el al. 2013). For the past three decades, BIM has been evolving to support data integration in the AEC industry (Cavka et al. 2017). BIM is ought to facilitate collaboration among different parties involved in a project, such as architects, engineers, consultants, contractors, facility managers, and owners. This collaboration is ought to be achieved via continual information exchange. According to Pauwels and Terkaj (2016),

1 PhD student, Teesside University, Middlesbrough, UK, y.alani@tees.ac.uk
2 Prof, Teesside University, Middlesbrough, UK, n.n.dawood@tees.ac.uk
3 Dr, Teesside University, Middlesbrough, UK, s.rodriguez@tees.ac.uk
4 Dr, Teesside University, Middlesbrough, UK, h.dawood@tees.ac.uk

the IFC standard that was developed by BuidingSMART was aimed at supporting data exchange via providing a central "conceptual schema and an exchange file format for BIM data" (IFC4 Documentation 2020). However, the interoperability issue still exists, and it is prominent in the literature. For instance, We et al. (2019) stated that current BIM object databases suffer from the lack of unified classification system of building components, and thus, negatively affecting interoperability. Moreover, Zhang et al. (2018) further supported this argument as they claimed that when working with IFC instance building models—which is the state of the art in the AEC industry—challenges are encountered by industry practitioners when retrieving domain specific information. They claimed that solutions are usually proprietary and vender specific. Also, Godager (2018) stated that to achieve BIM level 3 and beyond, it is important that data are searchable by both machines and humans—though BIM "maturity levels" are no longer promoted, ISO 19650 series—(UK BIM Framework 2020). Hence, product manufacturers are important actors in the process, yet their participation is constrained by the lack of dynamic links with other domains (Costa and Madrazo 2015). This paper explored establishing these dynamic links to facilitate the information flow from product manufacturers, all the way to the asset management software.

In this regard, one area of research that seems promising is Semantic Web Technologies (SWT). SWT are usually used to create ontological models and have become rather prominent in the AEC research community, particularly in the last decade. Zhong et al. (2019) conducted a scientometric analysis on the ontology research within the construction industry from 2007 to 2017. Their results showed that researchers were initially focused on using ontologies in different areas of construction management. However, they concluded that 2016 marked the year after which researchers' focus shifted towards solving interoperability issues across the building life cycle.

1.2 Semantic Web Technologies (SWT)

The main concept behind SWT is to provide and connect ontologies, enabling inference on the data level rather than on the schema level (Pauwels et al. 2017), which is expected to also provide smarter querying. Data models are governed by rules, and complicated models would usually have a large number of rules. Indeed, the more rules in a model, the more processing layers are required. That is not the case when modelling with RDF, however, as it was partially designed with this in mind (Allemang and Hendler 2011). The Resource Description Framework (RDF) is a Semantic Web modelling language that is recommended by the W3C (semantic web – W3C 2020). Databases that adopt the RDF format are usually referred to as triple stores. To represent a relational database in RDF, each row, column, and cell would be allocated an IRI, and hence, giving value to the datum itself. A piece of information represented in RDF is known as the subject-predicate-object triple, figure 1. The object of one triple would form the subject of another and so on. This would result in a family-tree-like data or graph data. Using these technologies, different knowledge domains are represented as ontologies. Gruber (1993) defined an ontology as an "explicit specification of a conceptualisation" and Borst (1997) added to this definition, defining an ontology as "explicit specification of a shared conceptualisation". Therefore, an ontology should be formal, sharable, extendable, and reusable by others. RDFS-Plus and the Web Ontology Language (OWL) are modelling languages built on top of RDF, that differ in their complexity, and hence, inferencing capability. According to Allemang and Hendler (2011), there is a trade-off between the complexity of the ontological model and the simplicity of the required queries. A complex ontology, that is a one that has many properties and restrictions, requires relatively simple queries to return the desired results. Whereas a simpler ontology would require longer queries to return the same results. Therefore, the model/ontology must be designed with querying in mind. Nonetheless, Allemang and Hendler (2011) also stated that increasing the model's complexity may make other applications associated with it more complex.

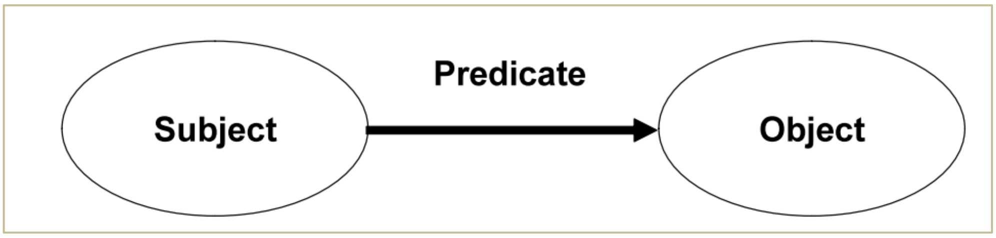

Figure 1: RDF triple

2 METHODOLOGY

This paper presents an industry-based research, in which, a major water company in the United Kingdom would act as the subject of the case study (data provider). The water company suffers from the lack of automation in asset data flow from the suppliers to the asset management software. The major issues that this project is ought to solve can be summarised in the following:

- Data modelling and querying.
- Automation of data flow from suppliers' databases to the asset management system.
- Asset management.

Therefore, a literature review was conducted to understand the trends in research on applying SWT in the AEC industry, verify the company's issue as a research problem, and understand the current state of the art and explore some of its current applications. Then, a framework has been developed to tackle these research issues. The methodology for building this framework was adopted from Dawood and Vukovic (2015), who generally classified WLC information flow into four pillars:

1. People
2. Process
3. Policy
4. Technology

The framework in this paper is aimed at the technology pillar, summarised in the flow chart in figure 2, which consists of seven steps. This framework is to investigate the potential of using SWT in solving the problem of data integration, using a WLC of information flow approach.

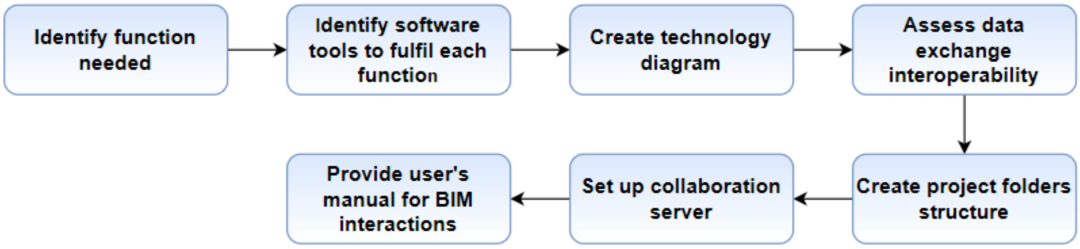

Figure 2: WLC information flow-technology pillar (Dawood and Vukovic 2015)

According to Stark (2011), "Product Lifecycle Management (PLM) is the business activity of managing, in the most effective way, a company's products all the way across their lifecycles; from the very first idea for a product all the way through until it is retired and disposed of". WLC of information flow can be thought of as a part of PLM for the construction industry. Dawood and Vukovic (2015) defined WLC as "the steady and continuous evolution and use of BIM information and knowledge from the design stage, through the construction stage, facility management (FM) stage, and up to decommissioning" achieved through rules or graphical process maps. On the other hand, a Building-Life Cycle (BLC) approach is defined as the integration and consistent evolution of BIM information throughout the project's life cycle (Underwood and Isikdag 2010). BLC consist of the initial planning and design of project, construction phase, operation and maintenance, and dismantling and recycling phase (Farias et al. 2018). As the project evolves from one phase to another, information loses value. For every succeeding stage, the information provided by the previous stage is usually missing, ambiguous, poorly structured, etc. (Dawood and Vukovic 2015). This was the main driver for a WLC approach; to reuse information and prevent its loss. WLC differs from BLC in that the former not only focuses on integration and coordination, but also the knowledge generated along the RIBA (Royal Institute of British Architects) Plan of Work Stages (RIBA 2020). This paper builds up on the research by Dawood and Vukovic (2015), focusing on the technology pillar of WLC, as they interviewed industry leaders in Qatar and compared their views with the UK's industry. In this regard, the findings of Dawood and Vukovic (2015) concluded that BIM technologies do not have major shortcomings. Yet, IFC exchange format caused data loss and geometry distortion during information exchange. BIM data is at the core of the technology pillar, which is why incorporating SWT with BIM is becoming a research attraction in AEC. Within the context of RIBA Plan of Work Stages (figure 3), this research looks into the detailed design stage, relevant to product procurement and asset management. RIBA Plan of Work in its latest update consists of eight stages. According to RIBA Plan of Work, it is essential to understand that highly detailed information will only start being delivered at stage 3 and upwards, as the client's requirements will become clear enough for a detailed design. Thus, it is at stage 3 when the level of input from the engineering team would need to accelerate. This paper is aimed at helping to deliver the objectives of stages 4 to 7. In stage 4 (i.e. technical design), all the design information required to manufacture/construct the project must be complete. Thus, when discussing construction stages, this paper will be referring to the RIBA Plan of Work.

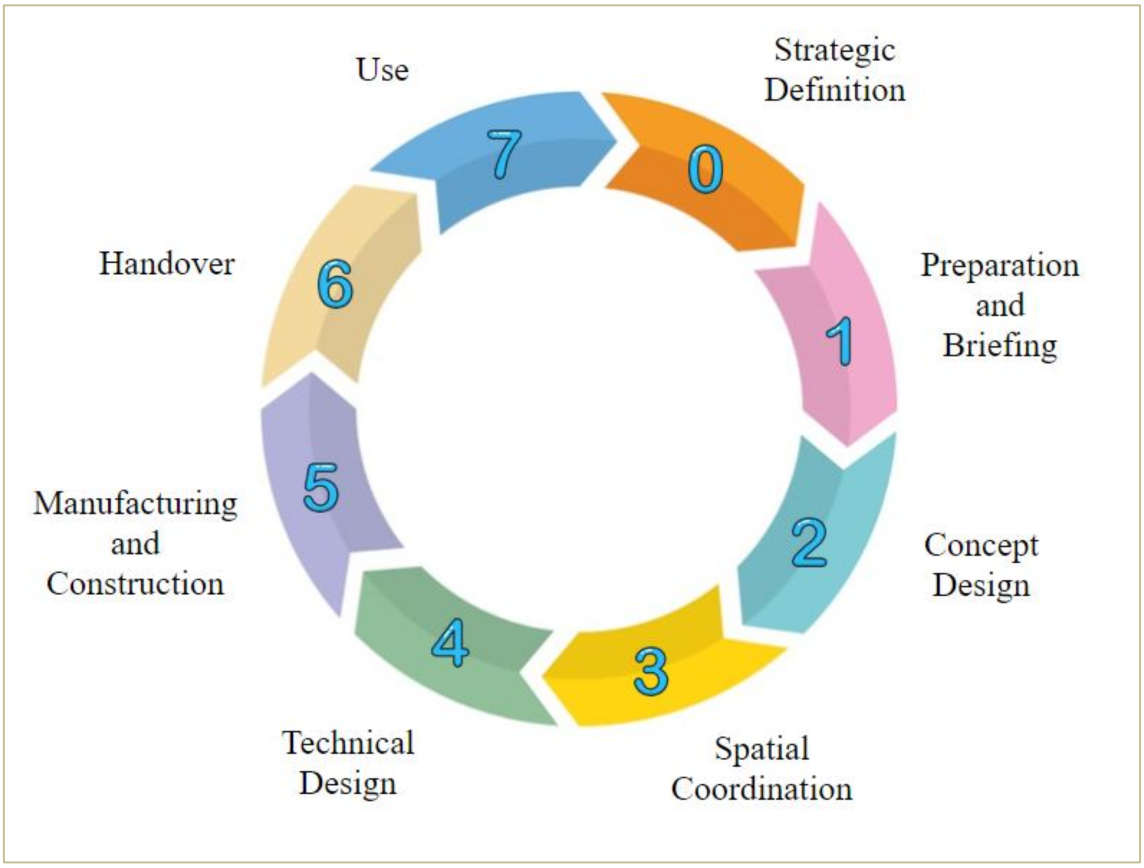

Figure 3: RIBA Plan of Work (RIBA 2020)

3 SWT IN AEC

Interoperability, linking across domains, as well as the logical inference and proofs are areas that SWT are expected to enhance within the AEC industry, according to Pauwels et al. (2017). They described interoperability among different domains as loading the same content into multiple applications, and linking across domains as combining different content available in multiple applications. Generally, SWT could allegedly offer improvements to data modelling in AEC, due to linked and continuously updated data. This signifies the importance of creation and maintenance of links between datasets, as Pauwels et al. (2017) argued that this would need human intervention often, resulting in the interoperability issue existing on a finer scale, i.e. at the data level. Nonetheless, SWT tackle data modelling at the finite data level, and thus they harness knowledge representation, which makes SWT of importance across all AEC domains. Abanda et al. (2013); Pauwels et al. (2017) and Zhong et al. (2019) have conducted exhaustive literature reviews on SWT in AEC industry, and the reader is encouraged to explore those resources. However, some of the recent practical applications of SWT in AEC will be presented here.

Rasmussen et al. (2019) created an Ontology for Property Management (OPM), for modelling complex properties in the design environment. They focused on answering competency questions when designing their ontology. Furthermore, they proposed an API to query their ontology and retrieve design data. In general, 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”. Kuster et al. (2020) created an ontology for urban district sustainability assessment (UDSA), in which, they incorporated Internet of Things (IoT) technologies. They demonstrated their work via querying the ontology for sensor data. They concluded that the model has proven helpful in achieving a linked data approach for urban district sustainability evaluation. Moreover, Simeone et al. (2019) stated that for built heritage buildings, the informative models are usually semantically poor (data for heritage buildings are usually unstructured, incomplete, or missing). Hence, they proposed a semantic-enrichment framework to enhance BIM models via ontologies. Schneirder et al. (2017) proposed semantically representing the control logic of Building Automated Systems (BAS). They stated that similar concepts can be applied with BIM and BMS (Building Management Systems). The ontological model served as a knowledge base for rule-based verification of control of control logic in BAS. They demonstrated their methodology on controlling an air handling unit (AHU). They assumed that such ontologies can be developed automatically using authoring tools. Yet, Wu et al. (2019) and Barbau et al. (2012) recommended manual development of ontologies to ensure quality. In the context of compliance checking, Fortineau et al. (2019) stated that in a PLM process, rule-related data must be transmitted from one stakeholder to another, which is problematic as these stakeholders would probably use different modelling paradigms. Thus, they investigated rule checking using Semantic Web Rule Language (SWRL) in ontologies to enhance PLM. They focused on integrating existing business rules into a product-centric information system. Kim et al. (2018) proposed a semantic web-based FM database which links BIM data (semantically) to historical work records. Their ontology integrated IFC BIM with FM database. Hence, during the operation and maintenance (O&M) phase, geometry and attributes information are integrated with FM through the Semantic Web. The problem they discussed lies in the data repository; the ability to store, merge, and retrieve heterogenous information. They stated that using object-oriented inferred spatial knowledge can improve work management information (e.g. resource and duration), space management, and energy monitoring. Kim et al. (2018) claimed that to operate BIM in FM using SW, facility managers should be trained in both BIM and SWT. Niknam and Karshenas (2017) created an ontology to represented building information. They suggested having a general, top layer, ontology that would be shared among different AEC domains by extending it to be detailed and specific. In this regard, IFC schema has been developed into an OWL version, namely ifcOWL (Pauwels and Terkaj 2016), which is a current buildingSMART standard (ifcOWL – buildingSMART Technical 2020). However, Zhang et al. (2018) argue that IFC data models are designed for data creation and exchange, but not tailored for querying task analysis. They also stated that IFC does not include the entire spectre of the AEC domain such as building requirements and regulations, product classifications, urban planning, and sensor networks. This paper presents a framework to investigate the benefits of SWT for data integration in infrastructure projects, using a WLC of information flow approach.

4 THE CONCEPTUAL FRAMEWORK

In simple terms, the framework is aimed at improving data flow from the suppliers' websites to the relevant asset management software via SWT. Also, SWT are expected to improve data querying and inferencing. The conceptual framework will be discussed in relation to the seven steps of the technology pillar of WLC information flow, figure 2, as well as the RIBA Plan of Work (RIBA 2020), figure 3.

4.1 Identify function needed

The paper scratches the surface of the capabilities of SWT. The Previous sections gave a mere introduction to help the reader understand the framework, which is ought to deliver a WLC of information flow for one of the United Kingdom's largest water companies. The current problem lies in the data management techniques; these techniques are further limited by the proprietary asset management software. The company defines its assets according to the Uniclass classification system (NBS 2015) and descriptions of each asset along with its important attributes are stored in Excel sheets as Product Data Templates (PDTs), see figure 4. These PDTs are popular among water companies in the UK (BIM4WATER 2020). The process of acquiring new information, filling in the PDTs, and introducing it to the asset management software is fully manual. According to the company, this manual process is due to the interoperability issue. Therefore, the prime objective of this framework is to achieve automation of this process. The first stage in the process is data acquisition, which is manifested in obtaining good quality product manufacturer data. In this regard, enhancing the process via smarter querying, and consequently, decision making, would allow users to search and choose the most suitable model for the design (Wu et al. 2019). It would be ideal if manufacturers or suppliers published their product data in the RDF format to allow for native querying, which is ought to be regulated by the ISO 23386 (ISO 2020). However, it may take some time into the future for the RDF format to be widely adopted on suppliers' websites. Alternatively, one solution may be to create a dictionary-like ontology consisting of classification clusters of all the related vocabulary for a specific domain. The domain expert (e.g. design engineer) would use this ontology to produce meaningful queries to search product libraries (Gao et al. 2015). Additionally, Geo et al. (2017) used this ontology that they developed to annotate BIM documents and index them. This indexing was done via algorithms to determine semantic relationships. Once an asset is selected from the library, it would then be imported into the BIM model. Though, a problem that usually arises in such systems—in addition to the technological barriers—is determining the attributes that are important for the asset management system or BIM. For example, figure 4 shows some of the attributes of a submersible pump that the water company needs to know for maintenance. Determining the important attributes of an asset is usually an iterative process with no clear methodology, due to the large number of variables involved. This bears the critical issue of clearly defining the requirements in the AEC industry (Alani el al. 2019).

Template Version	2.0			
Template Date	25 May 2018			
Suitability for Use	Final version			
Classification System	Uniclass 2015			
Classification Code	Pr_65_53_24_86			
Classification Title	Submersible pump unit			
Creator	[illegible]			
Template Custodian	[illegible]			
Information Category	Parameter Name (ISO15926 where applicable)	Value	Units	Notes
Manufacturer Data				
Specifications	Manufacturer	1	Text	
Specifications	Manufacturer Website	2	URL	
Specifications	Product Range	3	Text	
Specifications	Product Model Number	Test1	Text	or code
Specifications	CE Approval	4	Text	number, yes, no
Specifications	Approvals	5	Text	Eg ATEX, Reg 31(4)(b), WRAS
Specifications	Product Literature	6	URL	
Specifications	Features	7	Text	Free text to describe product
Construction Data - Pump				
Specifications	Casing Material	8	Text	Eg cast iron, stainless steel
Specifications	Outlet Connection Type	9	Text	Eg flange or spigot for hose connection
Specifications	Impeller Type	10	Text	Eg closed, semi-open, open, screw, vortex
Specifications	Impeller Material	11	Text	Eg cast iron, bronze, stainless steel
Specifications	Shaft Material	12	Text	Eg Stainless Steel
Specifications	Seal Type	13	Text	Eg Double mechanical
Specifications	Seal Material(s)	14	Text	Eg SiC vs. SiC
Specifications	Coating	15	Text	Eg Fusion bonded epoxy, 250mμ
Construction Data - Motor				
Specifications	Type	16	Text	Eg squirrel cage induction or permanent magnet
Specifications	Insulation Class	17	Text	Eg F, G, H, etc. in accordance with BS EN 60085
Specifications	Insulation Temperature Rise	18	Text	
Specifications	Method of Cooling	19	Text	Eg Air, proprietary coolant, pumped liquid
Specifications	Cable Length	20	m	
Specifications	Cable Type	21	Text	Eg BS EN 50525-2-21 (type H07RN-F) (or equivalent)
Specifications	Cable Connectors Fitted (Y/N)	22	Text	
Dimensional Data				
Specifications	Overall Length	23	mm	
Specifications	Overall Width	24	mm	
Specifications	Overall Height	25	mm	
Specifications	Gross Weight	26	kg	Equates to operating weight of unit
Specifications	Shipping Weight	27	kg	Equates to dry weight of unit plus packaging allowance
Specifications	Access Clearance Top	28	mm	Access required for maintenance of this item
Specifications	Access Clearance Bottom	29	mm	Access required for maintenance of this item
Specifications	Access Clearance Left	30	mm	Access required for maintenance of this item (as viewed from discharge)
Specifications	Access Clearance Right	31	mm	Access required for maintenance of this item (as viewed from discharge)
Specifications	Access Clearance Front	32	mm	Access required for maintenance of this item (as viewed from discharge)
Specifications	Access Clearance Rear	33	mm	Access required for maintenance of this item (as viewed from discharge)
Specifications	Outlet Connection Size (NB)	34	mm	Eg NB100 - refer to group
Specifications	Flange Rating (if applicable)	35	Text	Eg PN10 / PN16
Specifications	Impeller Diameter (Fitted)	36	mm	
Specifications	Impeller Diameter (Maximum)	37	mm	
Specifications	Manufacturers Drawing Reference No	38	Text	Eg Detail drawing reference No from the manufacturer
Performance Data				
Specifications	Lower Limit Rotational Speed	39	rpm	Eg Minimum operating speed
Specifications	Upper Limit Rotational Speed	40	rpm	Eg Maximum operating speed
Specifications	Flow Rate	41	l/s	Eg At max speed and BEP, based on SG1 - water

Figure 4: PDT example of submersible water pump

4.2 Identify software tools and create technology diagram

Identifying the relevant tools and creating the technology diagram are the second and third steps in a WLC approach, respectively. These steps have been conducted simultaneously, while consulting the literature. Figure 5 shows the proposed semantic framework (technology diagram) to achieve the objectives of this paper. The following steps explain the framework's concept and tools in reference to figure 5.

1. The process starts with the XLS sheets that are used to produce an asset specification ontology (ASO), from their schema. These are the Product Data Templates (PDTs) that the water company uses to describe asset data necessary for maintenance. The PDT for a submersible pump in figure 4 is an example. However, other assets have different criteria to describe them, and consequently, many attributes in these PDTs are not repetitive. There are few translational languages that could automate the process of creating ontologies from XLS sheets, such as TopBraid Composer (TopBraid Composer 2020), and XLWarp (Langegger and Woß 2009). However, as the data in the PDTs are not entirely structured, the process of converting them to OWL should be done manually.
2. The NeOn methodology was used to create the asset specification ontology, by reusing and reengineering non-ontological resources (Suárez-Figueroa et al.

2015). This ontology, which is being developed via Protégé, an open-source ontology editor (Protégé 2020) will consist of classes, attributes, and relations. OWL DL is the language of the new ontology, which includes all the OWL constructs, as opposed to OWL Lite or RDFS-Plus. Some of the OWL DL constructs that were utilised in the asset specification ontology were; owl:oneOf, owl:disjointWith, and cardinality restrictions (Web Ontology Language 2009). Hence, the ontology can be considered as heavyweight (Fürst and Trichet 2006). Figure 6 shows a portion of the developed ontology. As discussed in subsection 1.2, an ontology is a data model that contains three necessary characteristics; formal, explicit, and shared. The ontology in this research is formal and explicit in that it is machine readable and represents real assets (real phenomenon), whilst incorporating Uniclass naming (NBS 2015). Publishing the ontology will be at a later stage in the project.

3. In this step, the asset specification ontology (schema) is deposited in a triple store or graph database, such as RDF4J (Eclipse RDF4J 2020) and GraphDB (GraphDB 2020), respectively. They both store RDF data in a SPRAQL endpoint. Zhang et al. (2018) utilised SWT for improved data querying, as they claimed that current solutions are vender specific. However, even for Semantic Web applications this issue will exist. For example, the chosen data store would affect the process, as most of these databases are proprietary and often encompass different features—such as querying capabilities, once their licence has been purchased. Also, sharing ontologies is another problem that one encounters when working with real life application cases. Due to data protection policies, which is an important research area, full exploitation of Linked Data may be hindered.
4. SPARQL is the query language commonly used to query ontologies. The objective is to use SPARQL to query libraries for product data. As mentioned in subsection 1.2, the more complex the ontology is, the simpler the queries should be. Therefore, as recommended by Allemang and Hendler (2011), this ontology will be designed for querying, which is only reasonable considering the large amount of data in the PDTs. Also, queries are not only to be made on the ontology (via SPARQL Endpoint), but also on the internet for product data according to the parameters specified in the ontology schema. In this regard, Zhong et al. (2018) developed a framework for extending SPARQL queries as defined functions. This framework was to query IFC data, however. Product data, if not published in an RDF format, would need a data converter (see step 6) to RDF format to instantiate the ontology in the graph database.
5. This step in the framework requires identifying the data formats that water assets suppliers use to publish their data.
6. There exist some translational languages to convert data from relational format to RDF such as D2RQ (D2RQ 2020) and YARRRML (YARRRML 2020). As per Akinyemi et al. (2020), when querying across multiple sources over the internet, as in this case, then federated SPARQL queries are to be used. For such on-the-fly querying, Akinyemi et al. (2020) recommended using an Ontology-based Data Access (OBDA) approach. OBDA provides a user-friendly approach to query relational data bases using SPARQL queries (Xiao et al. 2018).

As mentioned earlier, the data providers for this case study rely on the Uniclass naming convention, and so do other large infrastructure companies in the United Kingdom. However, the naming provided by Uniclass may not be suitable for every company as people are not expected to familiar with the entire Uniclass naming convention. In addition to this, classifying assets in different languages introduces even more obscurity in managing these assets. An ontology approach to this matter is ought to form a solution. As explained in subsection 1.2, an ontology assigns an IRI to each resource (classes, properties, and instances) acting as a unique “fingerprint” to these resources. Two resources on the internet with different names but the same IRIs, are ought to be the same. As a matter of fact, the ontology would infer “owl:sameAs” property between them, indicating that they are the same entity. Therefore, it is suggested that Uniclass should introduce IRIs into their classification system to aid solving this ambiguity. In this regard, the Architecture, Engineering, and Construction industry can resemble the biomedical industry. The biomedical industry has resorted to utilising ontologies for unambiguous references to biological concepts, as they have created a wiki space for collecting scientific ontologies, namely the Open Biological and Biomedical Ontologies Foundry—OBO Foundry (Allemang and Hendler 2011). Therefore, perhaps the construction industry should also resort to a wiki space for collecting relevant ontologies, which could be named the Architectural, Engineering, and Construction Ontologies Foundry (AECOF).

This paper is mainly ought to present a framework to support an ongoing case study. The project is currently at the fourth step of the WLC approach, namely assessing data exchange interoperability, figure 2.

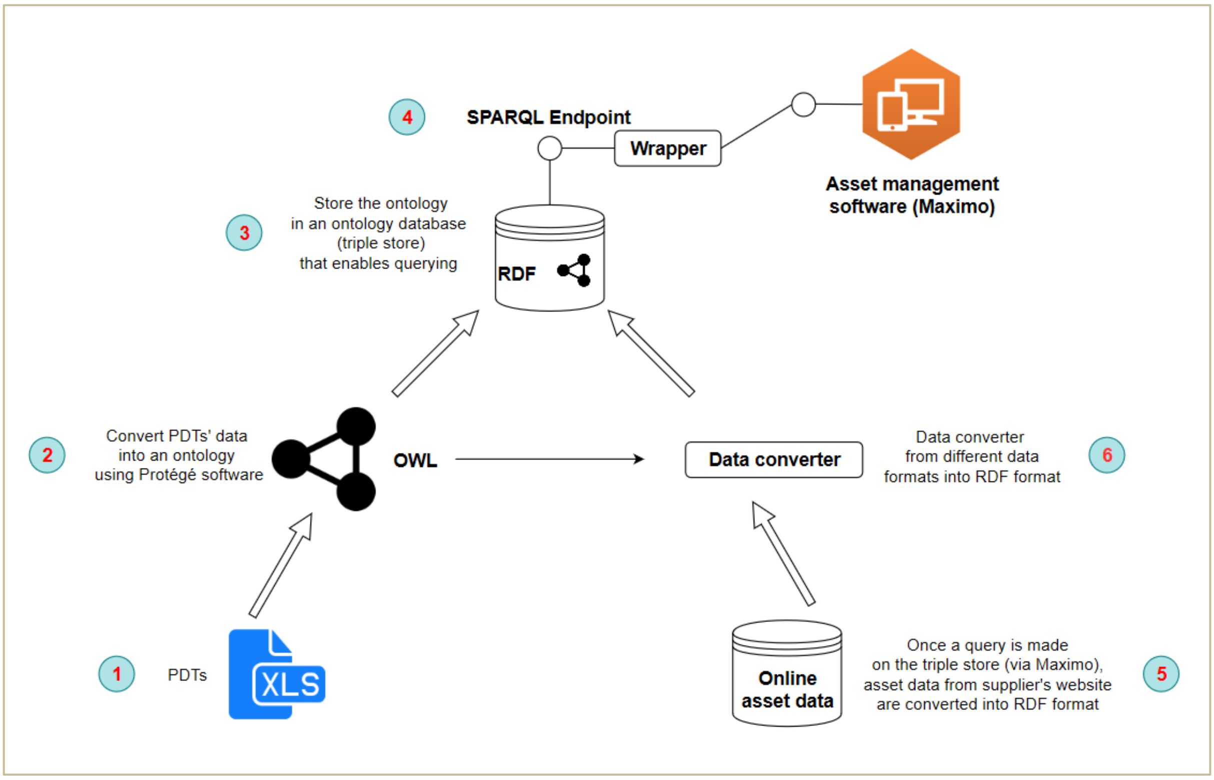

Figure 5: The proposed semantic WLC of information flow framework

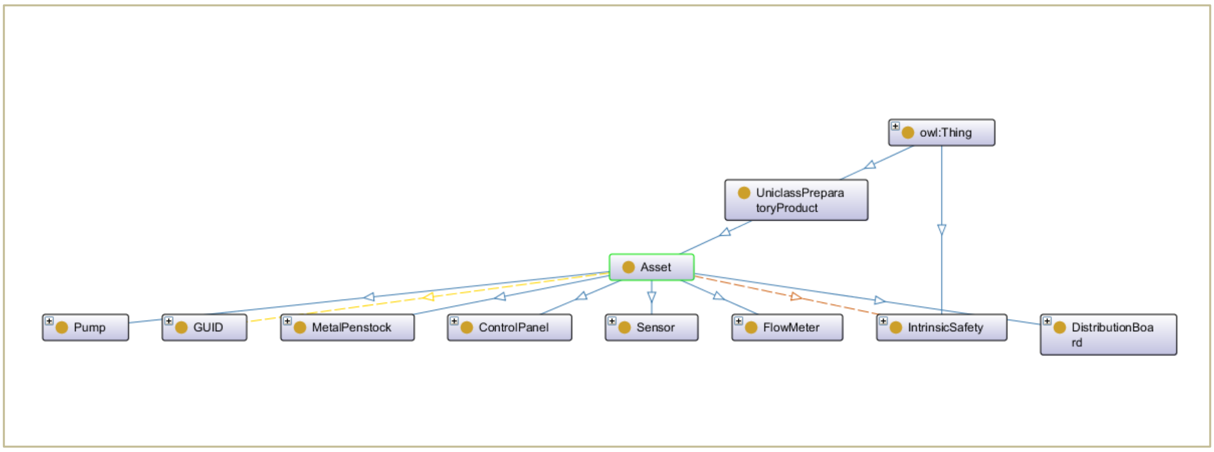

Figure 6: Portion of the water asset specification ontology

4.3 RIBA Plan of Work

As mentioned in section 3, the proposed framework is ought to aid the delivery of RIBA Plan of Work Stages 4 to 7 (RIBA 2020), figure 3. Stage 4 deals with the technical design as a responsibility of the design team and subcontractors. This paper proposes creating an asset specification ontology to be utilised at this point. An early asset specification ontology would encourage early facility management involvement, which is a current issue in AEC (Cavka et al. 2017; Alani et al. 2019). Thus, descriptive building system information may be preferred in this regard. Also, the proposed framework would aid the coordination between the design and specialist subcontractors' manufacturing information before the final design data are modelled in BIM. In the Manufacturing and construction stage, the construction team is usually the main actor. However, during this stage, the Semantic Web querying potential can be used on the asset specification ontology to verify design requirements. For example, spatial queries can be carried out via extending SPARQL queries as SPIN functions (Zhang et al. 2018). Similarly, in the Handover stage one of the outcomes is examining the final project pricing to issue the Final Certificate. Users can simply query an asset specification ontology to aid in determining the unit cost of all assets, contributing to the final pricing. The final stage is where most of the tangible benefits of this framework would be found. Most of the building's life cycle costs are during the operational phase (Kassem et al. 2015). It is assumed that the operational phase for utilities companies (infrastructure facilities) such as water companies, is expected to be even more significant in terms of maintenance than the building sector.

4.4 Limitations

Expected limitations to this framework include data availability, and data protection and ownership. If some data are not made available by the suppliers, then queries will not return desired results. Discussing the consequences of data protection and ownership is out of the scope of this paper. However, they form major constraints in the field of data sharing and integration.

5 CONCLUSIONS

Despite current state-of-the-art information exchange technologies such as IFC and ifcOWL, the AEC industry still suffers from inefficiencies of data flow. Particularly, the lack of product manufacturers' involvement in the process when dealing with infrastructure projects other than buildings. In this paper, a framework was proposed based on a WLC information flow approach to validate SWT as a solution to solve an industry-based (water infrastructure) problem, i.e. automating the process of information flow from product manufacturers to the asset management software. The literature seems promising in this regard, as some of the most recent application of SWT in AEC have proven successful. However, using federated SPARQL queries requires establishing mappings between the ontology schema and the data sources which is not a comprehensive solution. ISO 23386 (ISO 2020) is expected to bring benefits in this regard to help achieving Linked Data

One can still argue that research on application of SWT in the AEC industry is not intensively saturated, in comparison to research on BIM for instance. This claim is supported by Zhong et al. (2019) statement that research on SWT in the AEC have only started gaining researchers' interest in the last decade. This is further manifested in the lack of standards to support a semantic AEC industry. For instance, apart from ISO 15926 (Akinyemi et al., 2020), there were no ISO standards that support/mandate existing ontology deployment in the AEC industry, until the recent publication of ISO 23386 that promotes interconnected data dictionaries.

Data linking capabilities offered by SWT may be constrained due to proprietary licences and data protection policies. Ontologies are about having connected data in a shared environment; creating ontologies may be beneficial for querying, but as long as these ontologies are discrete, the full potential of SWT would be hindered. In general, researchers tend to construct their own ontologies to be tailored for their specific studies, and this paper is no exception. However, there needs to be more efforts on standardising and connecting good ontologies, in a similar manner to the OBO Foundry.

Future work is aimed at carrying out the case study with real data to test the framework, and validate the data using Shape Expressions (ShEx) or Shapes Constraint Language (SHACL) (Labra Gayo et al. 2018).

6 REFERENCES

Abanda, F.H., Tah, J.H.M. and Keivani, R. (2013), Trends in built environment Semantic Web applications: Where are we today?, *Expert Systems with Applications*, Elsevier Ltd, Vol. 40 No. 14, pp. 5563–5577, https://doi.org/10.1016/j.eswa.2013.04.027.

Akinyemi, A. G., Sun, M. and Gray, A. J. G. (2020) Data integration for offshore decommissioning waste management, *Automation in Construction*. Elsevier, 109 (May 2018), https://doi.org/10.1016/j.autcon.2019.103010.

Alani, Y., Hafeez, M., Patacas, J., Rodriguez, S. and Dawood, N. (2019). AN EXPLORATION OF THE OWNER'S REQUIREMENTS FOR BIM IN FACILITIES MANAGEMENT. In: *CIB World Building Congress*. Hong Kong, https://www.researchgate.net/publication/336881231_AN_EXPLORATION_OF_THE_OWNER'S_REQUIREMENTS_FOR_BIM_IN_FACILITIES_MANAGEMENT.

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

Barbau, R. et al. (2012) OntoSTEP: Enriching product model data using ontologies, *CAD Computer Aided Design*. Elsevier Ltd, 44(6), pp. 575–590, https://doi.org/10.1016/j.cad.2012.01.008.

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

Britishwater.co.uk. (2020). *BIM4WATER*. [online] Available at: <https://www.britishwater.co.uk/bim4water-standards-library.aspx#:~:text=Product%20Data%20Templates%20(PDTs)%20are,compliance%20with%20BIM%20Level%202.> [Accessed 25 January 2020].

buildingSMART Technical. (2020). Ifcowl - Buildingsmart Technical. [online] Available at: <https://technical.buildingsmart.org/standards/ifc/ifc-formats/ifcowl/> [Accessed 20 January 2020].

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, https://doi.org/10.1016/j.autcon.2017.08.006.

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, https://doi.org/10.1016/j.autcon.2015.05.007.

D2rq.org. (2020). D2RQ. [online] Available at: <http://d2rq.org/> [Accessed 21 January 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, https://www.researchgate.net/publication/275098377_WHOLE_LIFECYCLE_INFORMATION_FLOW_UNDERPINNED_BY_BIM_TECHNOLOGY_PROCESSES_POLICY_AND_PEOPLE.

El Ghosh, M., Naja, H., Abdulrab, H. and Khalil, M. (2017). Ontology Learning Process as a Bottom-up Strategy for Building Domain-specific Ontology from Legal Texts. In: *9th International Conference on Agents and Artificial Intelligence*, https://www.researchgate.net/publication/314521901_Ontology_Learning_Process_as_a_Bottom-up_Strategy_for_Building_Domain-specific_Ontology_from_Legal_Texts.

Eclipse rdf4j. (2020). Eclipse RDF4J. [online] Available at: <https://rdf4j.org/> [Accessed 21 January 2020].

Farias, T., Roxin, A. and Nicolle, C. (2018). A rule-based methodology to extract building model views. *Automation in Construction*, 92, pp.214-229, https://doi.org/10.1016/j.autcon.2018.03.035.

Fortineau, V., Paviot, T. and Lamouri, S. (2019), Automated business rules and requirements to enrich product-centric information, *Computers in Industry*, Elsevier B.V., Vol. 104, pp. 22–33, https://doi.org/10.1016/j.compind.2018.10.001.

Fürst, F. and Trichet, F. (2006). Heavyweight Ontology Engineering. In: *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*. [online] Montpellier: Springer. Available at: https://link.springer.com/book/10.1007/11915034.

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, https://doi.org/10.1016/j.aei.2015.10.003.

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, https://doi.org/10.1016/j.autcon.2015.04.006.

Graphdb.ontotext.com. (2020). Graphdb. [online] Available at: <http://graphdb.ontotext.com/documentation/free/about-graphdb.html> [Accessed 21 January 2020].

Gruber, T. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), pp.199-220, https://doi.org/10.1006/knac.1993.1008.

International Organization for Standardization. (2020). *Building information modelling and other digital processes used in construction – Methodology to describe, author and maintain properties in interconnected data dictionaries,* ISO 23386:2020

Kassem, M., Kelly, G., Dawood, N., Serginson, M. and Lockley, S. (2015). BIM in facilities management applications: a case study of a large university complex. *Built Environment Project and Asset Management*, 5(3), pp.261-277, DOI: 10.1108/BEPAM-02-2014-0011.

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, https://doi.org/10.1016/j.autcon.2017.12.019.

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, https://doi.org/10.1016/j.advengsoft.2019.102731.

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

Langegger, A. and Woß, W. (2009). XLWrap – Querying and Integrating Arbitrary Spreadsheets with SPARQL. In: *International Semantic Web Conference*. [online] Chantilly: Springer. Available at: https://link.springer.com/chapter/10.1007%2F978-3-642-04930-9_23

Li, H., Rezgui, Y. and Rana, O. (2013). Editorial for special issue: Cloud computing and distributed data management in the AEC – Architecture, Engineering and Construction industry. *Advanced Engineering Informatics*, 27(2), pp.158-159, DOI: 10.1016/j.aei.2013.04.002.

NBS. (2015). Uniclass. [online] Available at: <https://www.thenbs.com/our-tools/uniclass-2015> [Accessed 5 January 2020].

Niknam, M. and Karshenas, S. (2017), A shared ontology approach to semantic representation of BIM data, *Automation in Construction*, Elsevier B.V., Vol. 80, pp. 22–36, https://doi.org/10.1016/j.autcon.2017.03.013.

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, https://doi.org/10.1016/j.autcon.2015.12.003.

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, https://doi.org/10.1016/j.autcon.2016.10.003.

Protege.stanford.edu. (2020). protégé. [online] Available at: <https://protege.stanford.edu/> [Accessed 1 Febreuary 2020].

Rasmussen, M., Lefrançois, M., Pauwels, P., Hviid, C. and Karlshøj, J. (2019). Managing interrelated project information in AEC Knowledge Graphs. *Automation in Construction*, 108, https://doi.org/10.1016/j.autcon.2019.102956.

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].

Simeone, D., Cursi, S. and Acierno, M. (2019), BIM semantic-enrichment for built heritage representation, *Automation in Construction*, Elsevier, Vol. 97 No. November 2016, pp. 122–137, https://doi.org/10.1016/j.autcon.2018.11.004.

Schneider, G.F., Pauwels, P. and Steiger, S. (2017), "Ontology-Based Modeling of Control Logic in Building Automation Systems", *IEEE Transactions on Industrial Informatics*, Vol. 13 No. 6, pp. 3350–3360.

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

Stark, J. (2011). *Product Lifecycle Management* (Volume 1). 2nd ed. London: Springer.

Suárez-Figueroa, M., Gómez-Pérez, A. and Fernández-López, M., 2015. The NeOn Methodology framework: A scenario-based methodology for ontology development. *Applied Ontology*, 10(2), pp.107-145, DOI:10.3233/AO-150145.

TopQuadrant, Inc. (2020). Topbraid Composer. [online] Available at: <https://www.topquadrant.com/products/topbraid-composer/> [Accessed 19 February 2020].

UK BIM Framework. (2020). UK BIM Framework. [online] Available at: <https://ukbimframework.org/> [Accessed 10 February 2020].

Underwood, J. and Isikdag, U., 2010. *Handbook Of Research On Building Information Modeling And Construction Informatics*. Hershey, PA: Information Science Reference.

W3.org. (2009). Web Ontology Language. [online] Available at: <http://www.w3.org/TR/2004/REC-owl-features-20040210/#complexClassFull> [Accessed 6 February 2020].

W3.org. (2020). Semantic Web - W3C. [online] Available at: <https://www.w3.org/standards/semanticweb/> [Accessed 6 February 2020].

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, https://doi.org/10.1016/j.compind.2019.02.016.

Xiao, G., Calvanese, D., Kontchakov, R., Lembo, D., Poggi, A., Rosati, R. and Zakharyaschev, M. (2018). Ontology-Based Data Access: A Survey. In: *International Joint Conference on Artificial Intelligence (IJCAI)*. Stockholm: International Joint Conference on Artificial Intelligence, pp.5511-5519, https://www.inf.unibz.it/~calvanese/papers/xiao-etal-IJCAI-2018.pdf.

Rml.io. (2020). YARRRML. [online] Available at: <https://rml.io/yarrrml/> [Accessed 5 February 2020].

Zhang, C., Beetz, J. and de Vries, B. (2018). BimSPARQL: Domain-specific functional SPARQL extensions for querying RDF building data. *Semantic Web*, 9(6), pp.829-855, http://www.semantic-web-journal.net/content/bimsparql-domain-specific-functional-sparql-extensions-querying-rdf-building-data-0.

Zhong, B., Wu, H., Li, H., Sepasgozar, S., Luo, H. and He, L. (2019), A scientometric analysis and critical review of construction related ontology research, *Automation in Construction*, Elsevier, Vol. 101 No. November 2018, pp. 17–31, https://doi.org/10.1016/j.autcon.2018.12.013.