Data Integration Support for Offshore Decommissioning Waste Management

---

Abiodun Gideon Akinyemi

Submitted for the degree of

Doctor of Philosophy in Construction

on completion of research at

The School of Energy, Geoscience, Infrastructure and Society

August 2018

Abstract

Offshore oil and gas platforms have a design life of about 25 years whereas the techniques and tools used for managing their data are constantly evolving. Therefore, data captured about platforms during their lifetimes will be in varying forms. Additionally, due to the many stakeholders involved with a facility over its life cycle, information representation of its components varies. These challenges make data integration difficult. Over the years, data integration technology application in the oil and gas industry has focused on meeting the needs of asset life cycle stages other than decommissioning. This is the case because most assets are just reaching the end of their design lives. Currently, limited work has been done on integrating life cycle data for offshore decommissioning purposes, and reports by industry stakeholders underscore this need.

This thesis proposes a method for the integration of the common data types relevant in oil and gas decommissioning. The key features of the method are that it (i) ensures semantic homogeneity using knowledge representation languages (Semantic Web) and domain specific reference data (ISO 15926); and (ii) allows stakeholders to continue to use their current applications. Prototypes of the framework have been implemented using open source software applications and performance measures made. The work of this thesis has been motivated by the business case of reusing offshore decommissioning waste items. The framework developed is generic and can be applied whenever there is a need to integrate and query disparate data involving oil and gas assets. The prototypes presented show how the data management challenges associated with assessing the suitability of decommissioned offshore facility items for reuse can be addressed. The performance of the prototypes show that significant time and effort is saved compared to the state-of-the-art solution. The ability to do this effectively and efficiently during decommissioning will advance the oil and gas industry's transition toward a circular economy and help save on cost.

Dedication

This work is dedicated to my parents, Green and Olufunke.

Acknowledgements

First, I would like to acknowledge all the help that I received from my supervisors, Prof. Ming Sun and Dr. Alasdair Gray. Thank you for the guidance and encouragement during the period of my study.

Second, I would like to thank the management and staff of Ariosh Limited. The support in cash and kind is greatly appreciated. Also, my research was greatly influenced by the experiences I had on the projects I participated in while working for the company.

I would also like to thank my family for the love and support throughout the period of my study. Mum, I am thankful for the prayerful words. Dad, I thank you for taking care of my obligations while away from home. To my siblings and extended family, I thank you for the regular words of encouragement.

Lastly, special thanks to Dami Aguda for being a reliable flat mate, Prof. Paul Jowitt for being a great mentor and Heriot-Watt university for the scholarship and enabling environment to succeed.

Contents

List of Tables

List of Figures

List of Acronyms

AEC/FM	Architecture, Engineering and Construction/ Facility Management
BI	Business intelligence
BIM	Building Information Model
CAD	Computer Aided Design
CSV	Comma-Separated Values
DBMS	Database Management System
DECC	Department of Energy and Climate Change
DFD	Data Flow Diagram
DSS	Decision Support System
EDMS	Electronic Document Management System
EKP	Emergent Knowledge Process
EMS	Enterprise Management System
EIS	Executive Information Systems
ETL	Extract-Transform-Load
FN	False Negatives
FP	False Positives
GAV	Global as View
GLAV	Global Local as View
GUI	Graphical user interface
IFC	Industry Foundation Classes
IRI	International Resource Identifier

IS	Information Systems
Iso	Isometrics
JSON	JavaScript Object Notation
LAV	Local as view
MTO	Material Take Off
NIST	National Institute of Standards and Technology
OLAP	online analytical processing
OBDA	Ontology-Based Data Access
OWL	Web Ontology Language
P&ID	Piping and Instrumentation Diagrams
PCA	POS Caesar Association
PFD	Piping Flow Diagrams
PLL	Potential Loss of Life
RAD	Rapid Application Development
ROC	Receiver Operating Characteristic
RDL	Reference Data Library
RDB	relational database
RML	RDF Mapping Language
R2RML	Relational Database to RDF Mapping Language
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
RIF	Rule Interface Format
SQL	Structured Query Language
SPIN	SPARQL Inferencing Notation
STEP	Standard for the Exchange of Product model data
TN	True Negatives
TP	True Positives
UKCS	UK Continental Shelf
UML	Unified Modelling Language™
URI	Uniform Resource Identifier
URL	Uniform Resource Locator

WBS	Work Breakdown Structure
WWW	World Wide Web
W3C	World Wide Web Consortium
XML	Extensible Markup Language
ZWS	Zero Waste Scotland

Chapter 1

Introduction

1.1 Background

Data generated from various processes across the life cycle of an oil and gas facility are typically in varying formats and often required for activities other than those that generated them (Yang, 2009). This scenario is similar to the work process in the pharmaceutical (Gray *et al.*, 2014), construction (Wang *et al.*, 2014), and water management (Shao *et al.*, 2016) domains where data from multiple heterogeneous sources are required for answering key questions. According to Yang (2009), application of data from multiple sources in an integrated form is common in several activities across the life cycle of process industry assets. For example, operations personnel seek ways to optimise their work processes and improve safety by combining data from multiple sources (Devold, 2013); and at the decommissioning stage, the decommissioning option selection and waste materials management are activities that require integrated asset data (DNS, ZWS and ABB, 2015). These users and other stakeholders need the requisite up-to-date data to be available to support their decision-making process.

Offshore decommissioning particularly represents significant business opportunities for oil and gas service companies (RSA and ZWS, 2015). However, DNS et al. (2015) argue it is a liability for owners of offshore assets and regulators, because of the associated costs. Nonetheless, DNS et al. (2015) advanced the argument that a way of leveraging inherent business opportunities in offshore decommissioning is to mitigate its costs through the sales and reuse of decommissioned items. To achieve this effectively, reliable assessment of decommissioned items is required. Such an assessment relies on data collected on the various items over the life cycle of an offshore asset. Considering the contribution by Stacey et al. (2008) that offshore platforms have a design life of about 25 years and data management techniques and tools are constantly evolving, data captured about items to be decommissioned will be in varying forms. In addition, considering the many stakeholders involved with a facility over its life cycle, described by Yang (2009), information representation of the items will be heterogenous. These challenges make data integration difficult.

Chapter 1: Introduction

Over the years, data integration technology applications in the oil and gas industry have focused on meeting the needs of pre-decommissioning life cycle stages, such as construction management, maintenance planning and operations optimisation (Fiatech, 2011). This is the case because most assets are just reaching the end of their design lives (Royal Academy of Engineering, 2013). Currently, limited work has been done on applying the concept of data integration to legacy oil and gas data for offshore decommissioning waste management purpose (D3 Consulting Limited, 2017); and a report by industry stakeholders underscores this need (DNS, AMEC and ZWS, 2016). Consequently, this thesis considers how to integrate the common data types relevant in oil and gas asset decommissioning. The work is motivated by the requirements of steel and piping system components waste reuse, which require integrated life cycle data for reliability assessment. The methods developed are generic and can be applied whenever there is a need to integrate and query disparate data involving oil and gas assets.

1.2 Summary of Thesis

The aim of this thesis is to investigate the current decision-making environment in offshore decommissioning waste management and find ways of improving existing practices with the application of information technology.

To achieve this, the following objectives are identified:

i. *Analysis of the offshore decommissioning process*: the drivers, issues and opportunities of offshore decommissioning are explored, and its process decomposed to identify the tasks encumbered because of the lack of efficient data integration techniques and tools.
ii. *Analysis of existing data integration approaches: an analysis of existing data integration techniques and technologies is conducted, and a detailed comparison carried out.*
iii. *Design of a data integration system*: a system capable of integrating and querying distributed heterogeneous data sources is developed. The design ensures compliance to stakeholder requirements and the realities of the offshore decommissioning process.

iv. *Development of a life cycle data integration framework*: a formal model for life cycle data integration is developed. This will allow users to continue to use their existing practices and technologies. The model covers the integration and querying of distributed heterogeneous data.

v. *Validation of developed data integration framework with prototypes*: versions of the designed system are implemented with multiple use cases. Suitable existing technology platforms are adopted and built upon to achieve this. Performance tests are then conducted to investigate the efficiencies of the prototypes for comparison to existing practices.

This research is carried out within the context of project management in the oil and gas sector. As a result, the construction management benefits result from the Information Systems (IS) contributions. The contributions of this thesis include the following:

i. *Architecture for integrating heterogeneous data in distributed sources*: a key feature of this architecture is the ontology - a shared conceptualisation of a domain (Borgo and Masolo, 2009). This allows queries to be answered in an efficient manner by resolving the semantic obstacles between data sources.

ii. *Techniques that ensure that users can continue to use their existing relational database applications*: relational database is the traditional repository in the oil and gas industry (Djamaluddin *et al.*, 2018). Consequently, given the proposed solution is in a dissimilar technology platform, ways for extracting data from such repositories are required. This research recommended and demonstrated appropriate techniques to achieve this.

iii. *Reuse of project software resources*: this will improve productivity by saving on time and effort required for generating new solution artifacts for new tasks.

iv. *Prototype implementation of proposed data integration and querying systems for offshore decommissioning waste data management*: the prototypes implemented indicate recommended technology applications for achieving the proposed architecture. Implemented prototypes are used to demonstrate how decommissioned items can be assessed for reuse. This has potential to save on cost and benefit the environment.

1.3 Structure of Thesis

This thesis has 7 chapters and the following six are summarised as follows:

Chapter 2 covers literature on data integration and process industry. In the initial part, it introduces the basic concept of data integration, the main approaches, and the common solutions applied in the built environment sector. It shows that data integration allows independent heterogenous data sources to be accessed through a shared schema. In addition, it discusses the Semantic Web. In the second part, it describes the process industry elements including: the process features, stakeholders, infrastructure and life cycle stages. It also exposes the common types of data sources in the industry. Thereafter, it describes some of the life cycle activities that can significantly benefit from data integration and discusses the data quality dimensions that should be addressed to gain the benefits. In addition, the key challenges to data integration in the process industry were described, current practices used to address data integration explained, and the process industry data integration standard discussed. To conclude, the areas of decision analysis relevant to this study were discussed in detail.

Chapter 3 explains the offshore decommissioning process. Thereafter it describes the drivers, issues and opportunities of offshore decommissioning in the UK North Sea. It also describes a gap analysis carried out to investigate the application of data integration for offshore decommission waste management. From the analysis, it shows that the current practice for data integration is a manual process. It discussed this solution and argues for the application of ontology-based technique.

Chapter 4 describes the basis for the research methodology and details its design. Also, it explains all the techniques applied throughout the research.

Chapter 5 enumerates the requirements for the software solution. It matches the appropriate technologies with the proposed solution model. These technologies and coding requirements are thereafter described. The testing, modification and evaluation of the developed software are also discussed here.

Chapter 6 describes the application of the developed solution model to offshore decommissioning waste management. It demonstrates and evaluates two use cases on piping, equipment and steel sections in process facilities.

Chapter 7 provides overall conclusion on regarding the achievement of the research aim and objectives. In addition, it highlights the benefits and limitations of the research. To conclude, it makes recommendations for further development of the work.

1.4 Publications

This research work has been submitted for publication as follows:

- Akinyemi, A. G., Sun, M. and Gray, A. J. G. (2018) 'A Data Integration Framework for Offshore Decommissioning Waste Management', Automation in Construction. *(Submission undergoing peer review)*.
- Akinyemi, A. G., Sun, M. and Gray, A. J. G. (2018) 'An Ontology-Based Data Integration Framework for Construction Information Management', Proceedings of the Institution of Civil Engineers - Management, Procurement and Law, 171(13), pp. 111–115. doi: https://doi.org/10.1680/jmapl.17.00052.

Chapter 2

Literature Review

2.1 Data Integration

Data integration provides a unified view to data stored in multiple, autonomous, heterogenous sources (Doan, Halevy and Ives, 2012a). As organisations and individuals typically generate and manage data in ways that suit their purposes, large amounts of data that are not logically related exist. Data integration helps to address this problem by combining data from multiple sources into a single model that can be queried.

2.1.1 Integration versus Interoperability

Integration and interoperability do not mean the same thing but are closely related. In reference to applications, integration means that two different applications are made to work together seamlessly (Fiatech, 2011). For example, if APP1 and APP2 in Figure 2.1(A) are made to work together by having their respective owners collaborate to develop a custom solution – connector, that would be called *integration*. Interoperability, on the other hand, is associated with two or more applications that work together by virtue of each, independently, following an external solution (Rajabifard, 2010). For example, if APP4, APP5, APP6 and APP7 shown in Figure 2.1(B) can work with one another by using, for example, an industry-wide approved connector, that would be called *interoperability*. In a broader sense, a software suite with multiple applications as shown in Figure 2.1(A or B) can be imagined having its applications integrated if information created by one of its applications is available to the other applications. However, the applications in one suite will not be interoperable with applications in another suite, if they do not comply with an external solution that can enable this.

In reference to data, integration can be achieved by point-to-point mapping or using a common neutral format (Fiatech, 2011). In point-to-point mapping, shown in Figure 2.1(C), an application pulls data from another application and uses data maps for transforming the pulled data to the format of the receiving application before importing it. The imported data can be *integrated* with data existing in the recipient application and data between the two applications is *interoperable*. Data integration using a common neutral format requires converting data from given sources to a given neutral format using adapters (Figure 2.1(D)). An adapter helps to transform data between a neutral format and the format of a given source. Data combined from sources in neutral format can be described as *integrated*, and the exchange of data via the neutral format, between applications, refers to *data interoperability*.

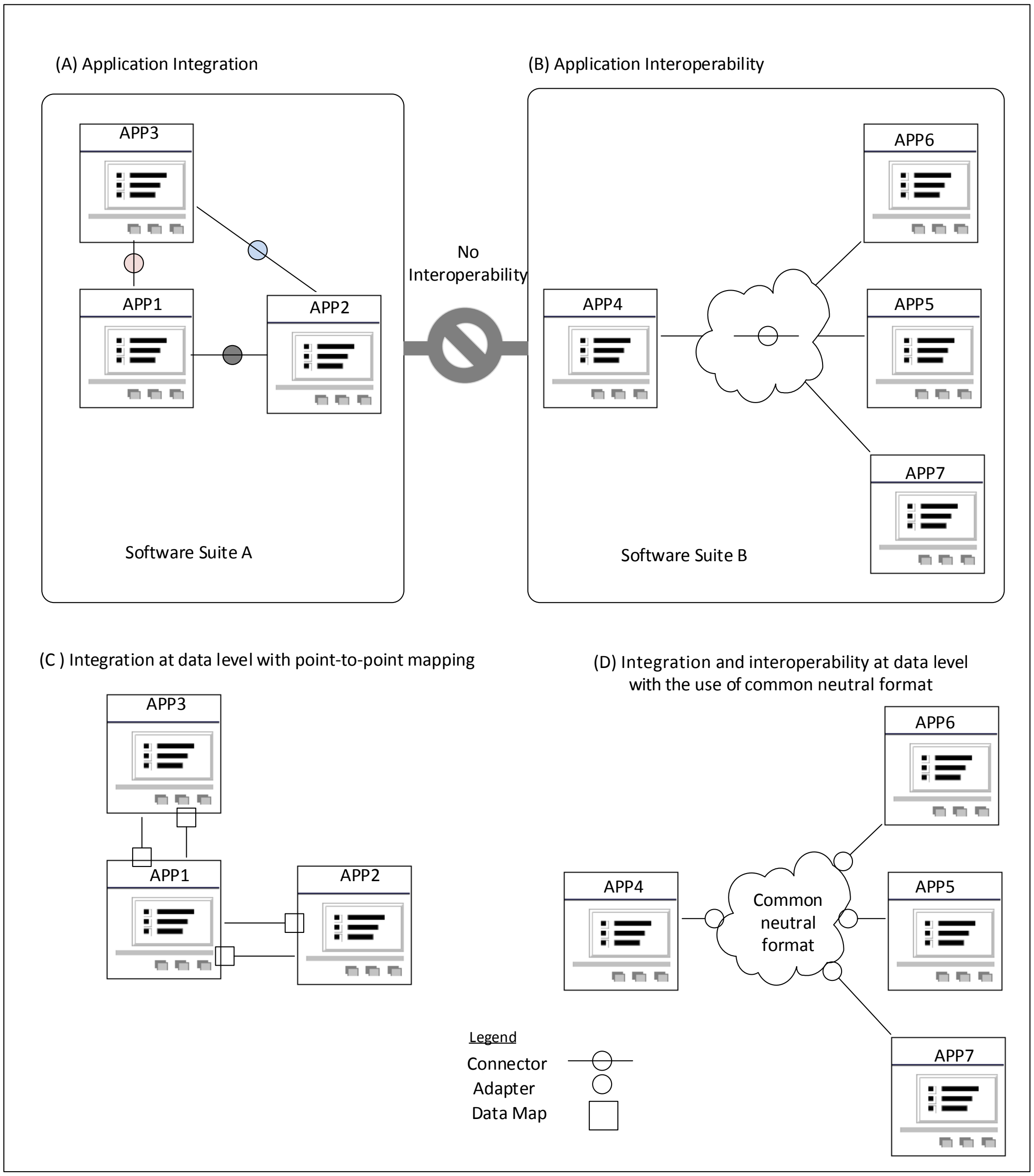


Figure 2.1: Integration versus interoperability

2.1.2 Data Integration Problems

The main challenge of data integration is making different systems talk seamlessly to one another(Doan, Halevy and Ives, 2012a). When machines communicate, the problem of implied meaning based on context becomes a serious barrier. Subtle differences in the meaning of terms and their use contexts become relevant. This scenario is analogous to a collaboration between individuals speaking different dialects of a language - they are likely to refer to similar things in different way (e.g. Americans calling petrol gas and the British calling it fuel). Also, their reference to a thing may have different contexts (e.g. Americans measuring length in feet and other countries measuring in metres). And where everyone in the collaboration is independent, speaking with the same dialect will not guarantee seamless working together, as the individuals may conceive things differently. The common data integration challenges can be viewed through the following prisms:

a) System architecture: in the real world, organisations rarely deploy information technology applications thinking of workflow challenges outside their domain, or uptake of future technologies. As a result, when it is time to query outside the domain or migrate data into a new technology, issues with data integration arise. Already, querying more than one data source at once is a challenge. Considering that the aspiration is to query as many sources as necessary for a search task, the challenge becomes even more complicated. In addition, considering the likely independence of each source (autonomy), automatic access is not a surety as the administrator of a data source can change the access credentials at any time without informing the users. This may be for privacy (medical, law enforcement, proprietary) (Clifton et al., 2004), or integrity (mission-critical) reasons (Doan, Halevy and Ives, 2012).

   Executing queries over multiple systems is challenging because the capabilities of each of the data sources in terms of the query processing power can be very different. For example, while one source may be a full SQL engine and therefore able to accept very complex queries, another source maybe a Web form and only able to accept a very small number of queries. Other sources may be non-editable documents, needing conversion and structuring before being made available for querying. Also, where there are differences in the ways vendors implement the query language, for example varying SQL implementations, these must be reconciled for data integration. Update of data is of interest in data integration because of the need to have the most current data. Where there are differences between the technologies of the source and target systems, updating data can become a challenge.

b) Data heterogeneity: because of the different requirements for individual tasks within an organisation, the creation of multiple heterogenous data products (e.g. files and databases) is inevitable. Heterogeneity of data can be structural (syntactic) or semantic (Wiesner, Morbach and Marquardt, 2011). These challenges have to do with the data models in the data sources, and the meanings attributed to their constituent entities. Because every information system is unique in these aspects, aligning data entities between information systems is always challenging and makes data integration difficult.

   Figure 2.2 describes some examples of data heterogeneity issues in data integration using structured data. The local schemas in the figure are assumed to be from different sources and the data they describe are to be integrated in the global schema. The source (local) and target (global) systems are assumed to be relational databases. So, depicted schemas specify tables with set of attributes. Figure 2.2(A) describes a semantic issue with table naming. 'Service' relation in the figure has relevant information for 'Maintenance' relation. However, different names used for the relations may affect easily matching the records. Another example of semantic issue is at the attribute level where 'label' in the 'Service' relation and 'dialNumber' in the 'Contact' relation have the same meaning as 'tag' and 'telephone' respectively in the 'Maintenance' relation. The use of different names for these attributes may confuse the record matching process. Figure 2.2(B) shows an example of syntactic issue related to the tabular organisation of schema. 'date' in the 'Service' relation is recorded as a single value. However, the value needs to be broken down into 'day', 'month' and 'year' attributes when integrated in 'Maintenance' relation. Another example of syntactic issue relates to schema coverage. This is described in Figure 2.2(B) where 'Service' and 'Contact' relations contain 'remark' and 'website' attributes which are not required in the 'Maintenance' relation.

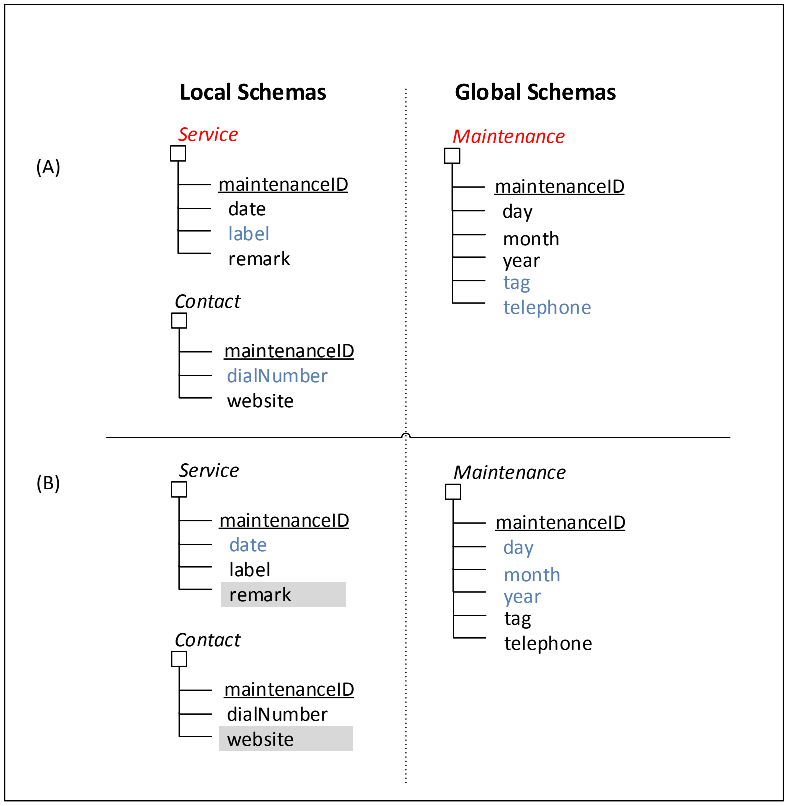


Figure 2.2: Data heterogeneity issues

2.1.3 Data Integration Approaches

The common approaches to data integration include:

a) Use of Middleware: this type of solution makes use of a custom-built software (middleware) that transfers information between two disparate applications using data maps (Fiatech, 2011). For each pair of applications, a unique software must be created. The software development involves a brute force approach that requires an enormous coding effort. This integration approach can be used to manage multiple applications. As the number of applications increases, the number of unique interfaces connected to middleware grows exponentially. For example, while two applications will have two interfaces, five applications will have 20 interfaces. The formula for calculating the number of interfaces in a middleware configured system is as given in equation 2.1 where n is the number of applications.

$$Number\ of\ Interfaces\ = n(n-1) \tag{2.1}$$

Considering the rapid pace at which new applications are being developed, the data structure of an application within a middleware configured system is bound to change. When this happens and there are point-to-point maps between the changed application and other applications, associated links will break. To fix the links, the data structures of the applications must be relearned. In cases where the internal data models are poorly documented, this will not be an easy task to accomplish. Rather, it will result in a high maintenance cost for this integration method (Wiesner, Morbach and Marquardt, 2011). The use of middleware for data integration is best applied as a short-term solution in systems with few applications.

b) Data Warehousing: this is also known as materialised database. In this approach, mappings are defined for loading data in the sources to the data warehouse schema. Data import is executed using Extract-Transform-Load (ETL) tools. A challenge with the approach is that data must be reloaded frequently to work with the most up-to- date version of source data. Advantages of the approach include its support for complex queries and efficient query processing (White, 2018). Figure 2.3 shows a generic architecture for data warehousing approach to data integration.

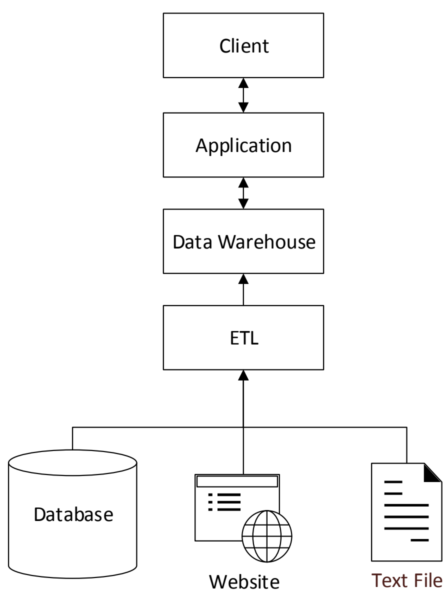


Figure 2.3: Exemplar architecture for data warehousing

c) Virtualisation: this makes use of a global schema. Mappings are defined for relating data in the sources to the global schema and user queries parsed over the latter, using the mediator, are rewritten over the sources through wrappers for distributed execution. Answers to queries posed to the global schema are generated by combining the results of sub-queries posed to the local sources. This approach to querying depends on the expressivity of the mappings between the mediator and the wrappers. An advantage of this method is that data is always fresh because queries are redirected to the original sources (Chamanara, König-Ries and Jagadish, 2017). Figure 2.4 show a generic architecture for virtualisation approach to data integration. The assumption is that each source is autonomous and with a unique schema. The sources can also be managed by one to many organisations. The three types of schema level mapping (Xu and Embley, 2004) used for virtual data integration are discussed as follows:

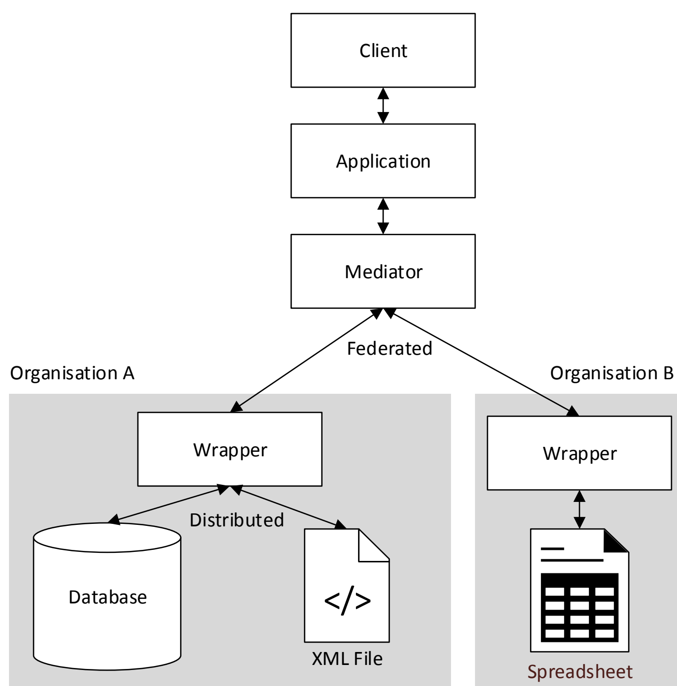


Figure 2.4: Exemplar architecture for virtual data integration

i. *Global as view* (GAV): in this method, tuples are expressible over the sources through the mediator. Queries are expressed in terms of the global schema and processed through their expansion to the sources. An advantage of this method is that the query writing process is straightforward. A challenge with

GAV is that every new source added to the data integration system will require changes to the existing mappings. This is because the global schema needs to model relevant source data to query it, and the mappings used for the source data need to explicitly specify how they are combined with other sources to form global relation tuples.

ii. *Local as view (LAV)*: in this approach, tuples are expressible over the global schema. Queries are expressed in terms of the local schema and only able to realise a subset of the data expressible in the global schema. Advantages of this method include the ability to add new sources without changing existing mappings between the sources and the global schema, and ability of the sources to accept constraints. These qualities are possible because the sources register independently in the global schema and the mappings of a source do not refer to other sources in the system. A challenge with LAV is that, just like GAV, it cannot query sources that do not have their data modelled in the global schema. Another is that, unlike GAV, it involves complex query writing process - queries posed to the global schema are answered by rewriting them as source views over the global schema.

iii. *Global local as view (GLAV)*: this method combines the earlier two approaches and tries to resolve their limitations. In this approach, queries are expressed over both the local and global schemas.

An example to illustrate the three types of mapping is presented in Figure 2.5. The global schemas shown in Figure 2.5(A) provide information on maintenance and its types. The 'Maintenance' relation represents identity number, date and label of maintenance service provided. Also, it represents the contact phone number for the service provider and the source of the record. The 'Maintenance_Type' relation represents maintenance identity number, and remark on maintenance service provided. The local schemata in in Figure 2.5(A) represent two data sources. One contains relation about maintenance service attributes; and the other, contact detail of maintenance service provider. The 'Service' relation stores the maintenance identity number, date, label and remark for services provided, while the 'Contact' relation stores the phone contact and website address related to a maintenance identity number.

Given the described relations in Figure 2.5(A), an example of GAV mappings is shown for 'Maintenance' relation in Figure 2.5(B). To relate the local relations to the global schema for this mapping technique, view in Equation 2.2 is produced. Note that "_" represents a variable that has been left out.

$$\textit{Maintenance (maintenanceID, date, label, phone, "Service")} \Leftarrow \textit{Service (maintenanceID, date, label, \_),} \\ \textit{Contact (maintenanceID, phone, \_)} \tag{2.2}$$

Again, given the described relations Figure 2.5(A), an example of LAV mappings is shown for the 'Service' relation in Figure 2.5(C). To relate the global schemas to the data source, view in Equation 2.3 is produced.

$$\textit{Service (maintenanceID, date, label, remark)} \Leftarrow \textit{Maintenance (maintenanceID, date, label, \_,"Service"),} \\ \textit{Maintenance\_Type (maintenanceID, remark)} \tag{2.3}$$

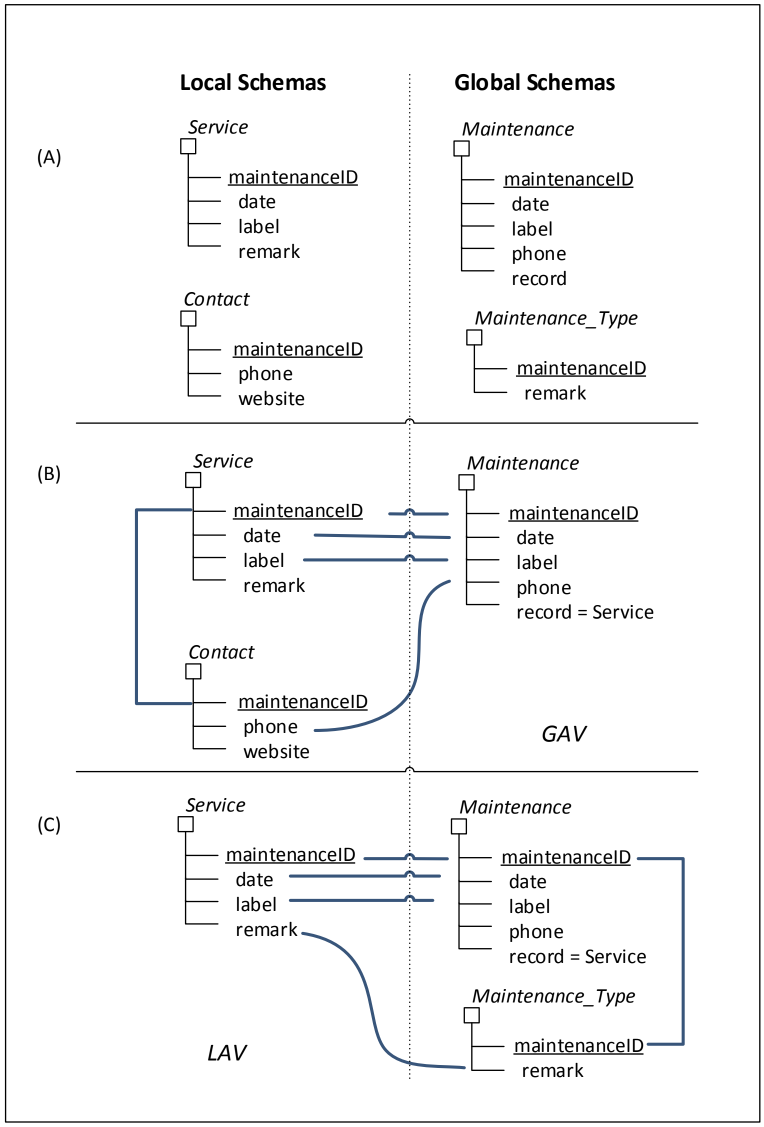


Figure 2.5: Examples of view-based data integration mapping

Lastly, using the described relations in Figure 2.5(A), an example of GLAV mappings relating the local relations to the global schema would be a combination of views in Equations 2.2 and 2.3.

2.1.4 Built environment solutions

According to Mutis and Issa (2012), integrating data from multiple sources in the built environment can be labour intensive and often incurs a high cost in terms of expert resources. This is because every software package has a unique way to structure data entities and label their properties – called the data model. The data model of an application may not be open, resulting in data integration difficulty. Although technologies like XML and relational database schemas can handle the syntactic differences between structured heterogeneous sources, these technologies cannot resolve the semantic incompatibilities without a careful mapping effort (Wiesner et al., 2011). The commonly used data integration techniques in the built environment sector include:

a) Document-based information integration: In this method of data integration, there is no knowledge of the content of the source files other than the metadata attributed to them. The metadata is also the basis for the storage of the sources in a database. Given that a lot of built asset data are available in propriety electronic formats, they can be easily managed in this way. However, this integration method is shallow and requires further exploration of the sources after document retrieval. An example is when all the documents related to a project are labelled with a unique code and stored in a database. This integration method will allow such documents to be retrieved with queries that connect them via the unique code. However, each retrieved document must be read with the appropriate software application to find out if it has relevant contents. Electronic Document Management Systems (EDMS) are an example of this technique (Qady and Kandil, 2012).
b) Data Mashup: In this method, the output data is not truly integrated as the source data are only juxtaposed for human interpretation. The function that executes this is coded into a software application with access to the data sources as shown in Figure 2.6(i). An example is the merging of 3D point cloud from laser scanning with 3D geometric model in the graphical user interface of an application for clash detection or design verification (Son, Bosché and Kim, 2015).
c) Schema-based integration: This type of data integration technique ingests data in its native format and reconciles it to the data model of the host application as explained in Doan et al. (2012c). As shown in Figure 2.6(ii), this process involves mapping data from multiple heterogeneous sources to the schema of a host application. The schema of the receiving application is the mediated schema. An

example of this in construction is combining Computer Aided Design (CAD) models from two different software applications by using a third application.

d) Model based Integration: This type of integration as described in Sun and Aouad (1999), is the usual configuration in Enterprise Management Systems (EMS) with a unified repository. The integration process has logical understanding of the transactions between the applications and services within the confederation; and allows full and logical data integration. For example, if there is a rule that application2 in Figure 2.6(iii) requires input from application1 for a task to proceed, then a missing input from application1 will freeze the process.

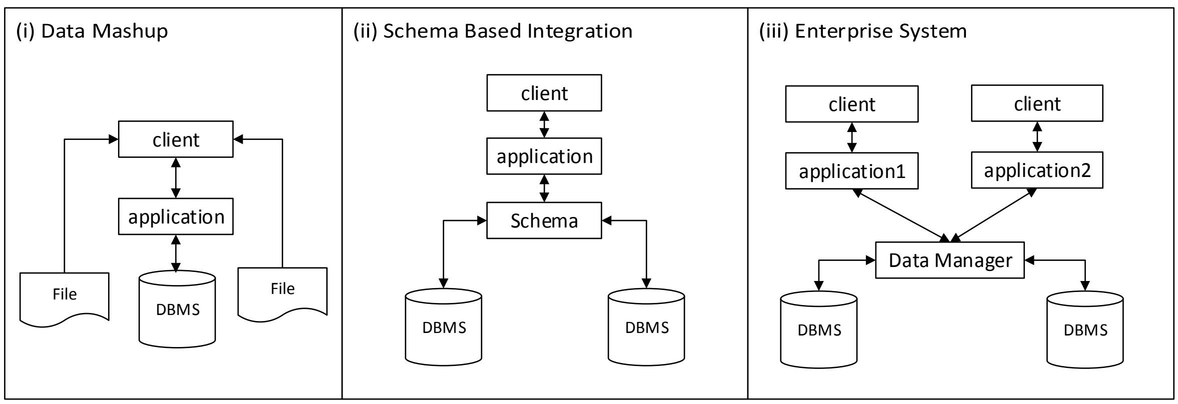


Figure 2.6: Illustration of common built environment data integration techniques

e) Ontology-based data integration:

i. *Ontology*: According to Studer et al. (1998),

"An ontology is a formal, explicit specification of a shared conceptualization. Conceptualization refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used, and the constraints on their use are explicitly defined. Formal refers to the fact that the ontology should be machine-readable. Shared reflects that notion that an ontology captures consensual knowledge; that is, it is not private of some individual but accepted by a group".

An ontology contains classes, relations, attributes, formal axioms, functions and instances. Classes are used to represent concepts that are either physical or abstract. When a class is derived from another class, it is called a subclass; and the class from which it is derived, called superclass. Relations are used to represent association between the concepts; attributes are used to describe the features of the concepts; formal axioms are used to make assertions about the ontology; functions are used for special cases of relations; and instances represent individuals within the ontology (Bermejo, 2007). Ontology types include (i) top level and (ii) domain specific ontologies. Top level ontologies use generic concepts for their entities (Borgo and Masolo, 2009) while domain specific ontologies make use of concepts that are common to users of a specific discipline e.g. oil and gas industry in Fiatech (2011). Ontology-based data integration makes use of ontologies in the integration of disparate data. The ontologies, according to Doan et al. (2012c), serve as the mediated schema and data sources are described in terms of their entities.

ii. *Semantic Web*: is the manifestation of Tim Berners-Lee's aspiration (Berners-Lee, Hendler and Lassila, 2001) of a web that is capable of automatic processing of data with minimal interaction with people. To achieve this, Doan et al. (2012c) explain that semantic markup is being associated with the content of the web to enable easier integration of heterogeneous data, and more accurate search results. To facilitate markup on the web, Resource Description Framework (RDF) (Cyganiak, Wood and Lanthaler, 2014), RDF Schema (RDFS) (Brickley and Guha, 2014) and Web Ontology Language (OWL)(Motik *et al.*, 2012) were developed. These languages according to Doan et al.(2012c) are based on the principles of knowledge representation languages which are grounded in description logic, a subset of first-order logic; and are used for defining ontologies.

Cyganiak, Wood and Lanthaler (2014) describes RDF as a directed, labelled graph made up of sets of triples – subject, predicate and object (see Figure 2.7C). It is a self-describing data representation language that supports several vocabularies at the same time and has several valid serializations (e.g. RDF/XML, Turtle and N-Triples) (Curé and Blin, 2015). RDF triples include International Resource Identifiers (IRIs), which provide a mechanism for referring to resources in a global way like the Uniform Resource Locators (URLs) used on the World Wide Web (WWW) (Fiatech, 2011). This will ensure that a resource is available to anyone wishing to use it. A concern with the use of IRIs stated in Doan et al. (2012c), is that they result in lengthy names for resources. Consequently, qualified names, which represent shorter names for the prefixes of resources, are used (see Figure 2.7A-B).

According to Brickley and Guha (2014), RDFS is used for modelling the classes, hierarchies and class membership of individuals in an ontology. Class refers to the nature of things that allows them to be grouped according to some criteria (ISO, 2003). RDFS is also used for specifying the domain and range of relationships in an ontology. All the constructs of the RDFS are contained in OWL. OWL in addition, has several constructs that improve its expressive power, enabling it to model more complex and incomplete domains. Doan et al. (2012c) states that OWL was developed to facilitate inferencing over the Semantic Web, and Motik et al. (2012) attribute the trade-off that exists between its expressive power and ability to perform inference for its multiple profiles.

The Semantic Web allows rules to be applied to ontologies modelled with RDFS and OWL (Horrocks and Patel-Schneider, 2011). These rules are defined using Rule Interface Format (RIF) (Bruijn and Welty, 2013) and they enable the discovery of extra information in the ontologies by generating new relationships based on existing ones. Concisely, the relevance of Semantic Web in ontology-based data integration includes: its use for developing the ontologies used as mediated schema; its use in describing and linking data from heterogenous sources, making them machine readable; and lastly, its support for reasoning about the integrated data.

iii. *Query Answering*: Having semantically enriched, searchable data can significantly improve the process of information extraction (Shayeganfar *et al.*, 2009). RDF data is queried using SPARQL, a World Wide Web Consortium (W3C) standard (The W3C SPARQL Working Group, 2013). SPARQL queries RDF data by matching the graph patterns between the queries and the data.

Considering the RDF triples in Figure 2.7C, the SPARQL query in Figure 2.7D selects the identity and designation of the items with location BD. It does this by matching the variable ?id in two triple statements. The result is as shown in Figure 2.7E.

(A)

Qualified Name	URI
basens:	http:/steel.waste.com/project/
qn1:	http:/steel.waste.com/project/rdl/
qn3:	http://steel.waste.com/properties/
pcardl:	http://posccaesar.org/rdl/

(B)

Resource Qualified Name	Resource URI
basens:S.03	http:/steel.waste.com/project/S.03
qn1:BD	http:/steel.waste.com/project/rdl/BD
qn3:hasLocation	http://steel.waste.com/properties/hasLocation
Pcardl:hasDesignation	http://posccaesar.org/rdl/hasDesignation

(C) RDF Triple

Subject	*Predicate*	*Object*
basens:S.03	qn3:hasLocation	qn1:BD
basens:S.07	qn3:hasLocation	qn1:CD
basens:S.07	pcardl:hasDesignation	“Cellar Deck”
basens:S.03	pcardl:hasDesignation	“Boat Deck”

(D) SPARQL Query

```
SELECT ?id ?deck
WHERE {
?id  qn3:hasLocation  qn1:BD.
?id  Pcardl:hasDesignation  ?deck
}
```

(E) Result

?id	?deck
Basens:S.03	“Boat Deck”

Figure 2.7: Example of RDF data and SPARQL query

2.2 Process Industry

The process industry is a part of the built environment concerned with processing bulk resources into products, and process plants are used for the various processes required to achieve these. Examples of process plant include chemical plants, oil refinery, gas plant, food processing facility, winery, brewery, pharmaceutical, Biotech facilities, nuclear plants, alternative energy plants, ammunition plants, plants for decommissioning chemical weapons, wastewater and water treatment plants, electric power plants, coal power plants, oil and gas, and mining plants (Yang, 2009). The following are some of the key elements related to process plants:

a) Process features: Process plants can be government or private owned and their processes can be proprietary or licensed. Generally, process plants are expensive to build, operate, maintain and decommission. Consequently, cost saving measures are always desired. Process parameters include process reactions, streams, composition, temperature, pressure and flow rates. These are contained in a Piping Flow Diagrams (PFDs) and Piping and Instrumentation Diagrams (P&IDs). PFDs show the relationship between the major components of a plant at a high-level while P&IDs depict them in detail. On a process plant there are instruments to measure process performance and ensure safety and piping is the mechanical system for the conveyance of process fluids between the process plant components. Pipings are made from different materials e.g. carbon steel and stainless steel. Also, they come in varying sizes and have different types of coatings. On a process plant there are various types of equipment. These include separators, reactors, vessels and pumps. Some of these are custom-made while others are available in standard market specifications.
b) Support Infrastructure: For a process plant to run properly, the supporting infrastructure must be adequate. This includes the road network, the utilities (e.g. electricity, water and sewage connections) and buildings as in the case of some process plants e.g. pharmaceuticals.
c) Life cycle stages:
    i. *Front-end Engineering*: this stage is about requirements gathering, scope definition, design definition, and evaluation of solution options. Deliverables like product specification, performance requirements, cost estimates - for development and operations, and development timeline are worked on in this phase.
    ii. *Detailed Engineering*: in this phase the various engineering disciplines involved with the design of a process plant work together in an integrated fashion to generate all the technical information required for the plant’s performance and configuration.

iii. *Procurement*: this is the phase where the equipment, components and materials specified in the engineering designs are acquired. Procurement phase overlaps the engineering and construction phases. Consequently, the three phases are usually handled by a single contractor.

iv. *Construction*: the items acquired in the procurement phase are used in the construction of the plant, ensuring compliance to the safety and design specifications recommended in the engineering designs. At the end of construction and before the handover to the plant owner or operator, mechanical completion to link up all the components of the plant system takes place. The results from testing of the plant system are normally required for regulatory compliance.

v. *Operations and Maintenance*: this is carried out by the operator or owner of the plant asset. It is the stage where the plant starts to produce the products for which it was built. During this phase, maintenance activities are carried out to keep the plant in operation and modifications are made to the plant system for efficiency gains.

vi. *Decommissioning*: this refers to the set of activities carried out at the end of life of a process plant to put it out of service. Activities involved include disconnection of the components, waste management and site remediation.

d) Stakeholders: throughout the life cycle of a process plant multiple stakeholders are involved. In the engineering and construction phases, the different disciplines of engineering (civil, mechanical, electrical etc.) are involved with the designs, and multiple contractors help to build different aspects of the project. In the operations and maintenance phase the plant operator continues to engage the various engineering disciplines and contractors for efficient running of the plant. During the life time of a plant, its ownership may change. This will introduce another set of stakeholders. Lastly to decommission the plant, the owner engages specialist decommissioning contractors, which are a different set of stakeholders.

2.2.1 Information Landscape

The key data sources in process plants as they exist today include:

a) Paper Document: technical information was recorded, distributed, transferred and stored on paper documents before computer programmes and databases came along. Standards, datasheets, design calculations etc. continue to exist on paper documents. Paper documents are subject to the interpretation of the reader where there are no guidelines on how it should be used. Once produced, they are difficult to change. Paper also deteriorates over time and is cumbersome to distribute. Finding specific information is difficult and indexing requires significant effort. Paper documents are not best suited for the highly dynamic process industry environment existing today. Nonetheless, they are still very much in use in several long-established process plants for practical reasons. They are usually required as legal instrument for approvals and as evidence of items delivery.

b) Electronic Document: they are a product of advancement in computers. They are equivalent to paper documents but electronically generated. Electronic documents are significantly more useful than paper documents because they have the capability for word search and their metadata can be indexed in a database (see Section 2.1.4a). The use of electronic documents for plant engineering development and management is a better scenario compared to the use of paper documents. This is because electronic files can be easily distributed, and modifications to their contents can be easily done e.g. spreadsheet documents and CAD drawings. Easy modification can however lead to multiple versions of a document, resulting in consistency issues. Another benefit of electronic documents is that data can also be extracted from them and organised as databases. Electronic documents are gradually replacing paper documents and are the dominant information type in the process industry (Yang, 2009). Paper documents, where possible, are scanned and converted to electronic documents (Sun and Aouad, 1999).

c) Relational Data: relational database is the traditional repository in the oil and gas industry (Robert Stackowiak, Venu Mantha, Alan Manewitz, 2015; Djamaluddin *et al.*, 2018) and it stores data using a relation model - in which data is expressed as tuples. This type of data can be easily retrieved, manipulated, distributed, stored and used to support multiple activities. A relational database can be used to generate documents e.g. a piping isometric drawing could be generated from data stored in an application's database. Also, from the same database, instrument datasheets, valve list, equipment specification etc. could be generated (Fiatech, 2011). Relational data also supports data warehousing where multiple databases are linked to a central repository. The challenge with relational databases is that there is no semantic reconciliation of data and users must specify the exact data requirements (Doan, Halevy and Ives, 2012b). Relational data allows for the extensive use of tags as a data approach in process plant engineering (Yang, 2009). A tag is basically an intelligent data object with attributes that can be organised into a relational database. Tags help to link documents, databases and 3D CAD models.

d) CAD Models: this type of data represents the geometric properties of a physical asset in 3 dimensions. 3D models are useful for operator training, facility walk through, constructability reviews, finite element analysis, configuration management (clash detection), construction sequencing, progress reporting etc. (Volk, Stengel and Schultmann, 2014). Yang (2009) describes an intelligent 3D model of a plant as its digital twin as entities in the model have the actual attributes of the physical replica.

2.2.2 Data Integration Needs

Process plant projects can be global with several teams and systems exchanging and combining information (Fiatech, PCA and USPI, 2015). Also, there is usually information handover between phases of a project or asset life cycle that requires data integration (Olubunmi and Ward, 2017). As a result, there is a need for data integration across the life cycle of process plant assets (Fiatech, PCA and USPI, 2015). Examples of activities requiring data integration are given as follows:

a) Engineering: in this phase, requirements analysis to fine-tune project goals needs data to be integrated from requirements documents, schedule tags, process documents, cost spreadsheets etc. Also, engineering designs require data integration because data is contributed from the different disciplines that are involved. Examples of data integrated during engineering design include 3D model, schedule tags, equipment tags and change requests.

b) Procurement: during bid management, data from design team and vendors e.g. items list, price catalogues, design criteria, etc. are integrated in generating requests for proposals. The winning proposal is integrated with schedule tags, purchase orders, payment invoices etc. during contract management.

c) Construction: in this phase, the planning effort involves constructability reviews which integrate 3D model data and schedule tags. Construction execution and monitoring involves extensive use of data tags which are integrated for fabrication, scheduling, material processing, quality control etc. At the commissioning of a process plant, engineering tags, 3D models, job records etc. are integrated for the systemization process, and operational tests are validated by reviewing data integrated from job records, design criteria database, acceptance tags etc.

d) Operations and maintenance: during this phase, activities such as change management, configuration management, cost control, safety planning and process optimization involve integrating data from multiple sources.

e) Decommissioning: at the end of life of a facility, relevant life cycle data are integrated for decommissioning activities including waste characterization and decommissioning options assessment.

The US National Institute of Standards and Technology (NIST) estimates the cost of inadequate interoperability in the US capital facilities industry to be $15.8billion per year (Gallaher *et al.*, 2004). This is in part due to data integration challenges. Consequently, addressing the problem is one way of saving on cost in the development and management of process plants.

2.2.3 Information Challenges

Built asset development and management is a collaborative endeavour involving multi-disciplinary teams including client, architects, engineers, consultants, contractors, etc (Niknam and Karshenas, 2015). Each member of these teams is responsible for some aspects of the project and often relies on information produced by others. Fiatech (2011) suggests that where all plant data are supplied in a single technology supplier system, the process of combining data and extracting information is somewhat straightforward. However, because of the multi-stakeholder nature of the process industry some flexibility is required with technology use to accommodate the domain requirements of each stakeholder (Omar and Nehdi, 2016). Consequently, relevant data for a task sometimes do not combine as required - because of data quality issues; and often need reprocessing to facilitate integration (Soibelman *et al.*, 2008). To maximise the value from process plant data, its qualities in relation to relevant tasks must be taken into consideration. The key data quality dimensions discussed in Westin and Sein (2014), and Bilal, Oyedele, Qadir, et al. (2016) are described as follows:

a) Volume: no one set of built asset data is likely to be large enough to become unmanageable. However, the total amount of data generated over the asset life cycle can be large enough that it can become challenging to sieve out useful information (Yang, 2009). For example, Fiatech (2011) describes handover process of a process plant as arduous and prone to error as data size grows because of the difficulty in aligning and making connections between multiple data sources.
b) Variety: built asset activities generate significant amounts of data in a wide variety of formats (Sun and Aouad, 1999; Mutis and Issa, 2012). Text documents, videos and pictures are usually plentiful (Soibelman *et al.*, 2008). Also, the various software applications used produce files in a variety of binary formats e.g. Extensible Markup Language (XML) in Yurchyshyna and Zarli (2009), and comma-separated values (CSV) in Bilal, Oyedele, Akinade, et al. (2016). Variety also relates to the variations that exist when multiple sources represent the same real-world entity differently, regardless of identical export formats.
c) Veracity: this is about information integrity and focuses on the accuracy and currency of data (Kitchin, 2014). It is important that data being used for tasks represents what they are intended to represent. When they do not, errors are introduced into tasks. Simple things like using an old report for a decision or planning based on inadequate drawing detail can have significant risk and cost implications. Veracity also covers managing null values, misleading values, outliers and non-standardised values that may be present in data (Rubin, 2014).
d) Velocity: this is about the need for regular updates to project data - underscored by Omar and Nehdi (2016). Progress reports and schedule data fall in this category. As there is always the need to have the most accurate and up-to-date information

in a timely manner, the frequency of updates requires consideration in managing process plant data.

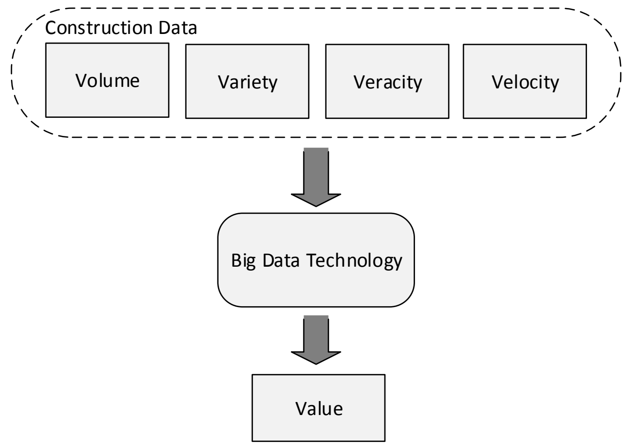


Figure 2.8: Extracting value from process plant data with big data technology

The highlighted quality dimensions of process plants data, as shown in Figure 2.8, are a replica of those associated with 'Big Data' by Kitchin (2014) except for the issue of very large data volumes – because typical asset data are manageable. Regarding data volume, Westin and Sein (2014) indicated that irrespective of the size of the organisation, investments must be made in information technology resources to ensure efficient handling of very large volumes of data. Consequently, issues other than data size are pertinent and Big Data technologies seem appropriate in addressing information challenges in related to highlighted data quality dimensions. Applying Big Data technology for process plant data management could mean investment in proprietary Big Data tools and skilled personnel, or developing own solutions using open source software.

Regarding data integration, information challenges associated with process plants are twofold. A part concerns legacy data, and the other, syntactic and semantic heterogeneity. The common formats of process plant legacy data include technical drawings, data sheets and manuals – which cannot be applied directly for data integration tasks; and described by Wang et al. (2014) as error prone and time consuming to extract information from. This may be because they are paper-based, unstructured, or versions of the authoring tools used in generating them no longer exist. Parts of typical facility information upgrade programmes focus on abstracting useful data from legacy sources and generating them as required using specialised tools (Olubunmi and Ward, 2017).

The second part, which is the focus of this research, deals with the issues hindering integration of heterogeneous data. This is premised on the fact that over the life cycle of an asset, multiple stakeholders generate information about it, and are likely to represent information about its objects and activities differently. This is because the data models of authoring tools are usually different (Bayer and Marquardt, 2004; Tolk and Diallo, 2005), and the domain understanding of some terms, according to Fiatech (2011), may be unique for different groups. As a result, attempts to combine data from multiple heterogeneous sources is not a straight forward process (Embury *et al.*, 2001). Meanings of terms must be resolved and efficient ways to handle large volumes of data in this form, effectively and cheaply, are required.

2.2.4 Common Data Integration Techniques

The two commonest data integration approaches in the oil and gas industry are the use of proprietary solutions, which require unique configuration by expert information technology personnel, or the manual approach using spreadsheets and databases.

a) Proprietary method: a typical approach for process plant data integration using custom tools is as shown in Figure 2.9 (Stidolph, 2015). The existing asset data is usually processed and organised into database systems. Documents go into document management systems and data is applied in some type of enterprise management system. This type of system is based on the data warehousing approach discussed in Section 2.1.3(b). Data that does not fit in any of the latter systems is stored away for future reference or until its retention period elapses. The end product of this type of effort is usually not a fully integrated information system, but a mix of database systems. Although, this may be what is practicable for an organisation’s work process, it usually brings about data integration challenges whenever there are new requirements that need data across the sources.

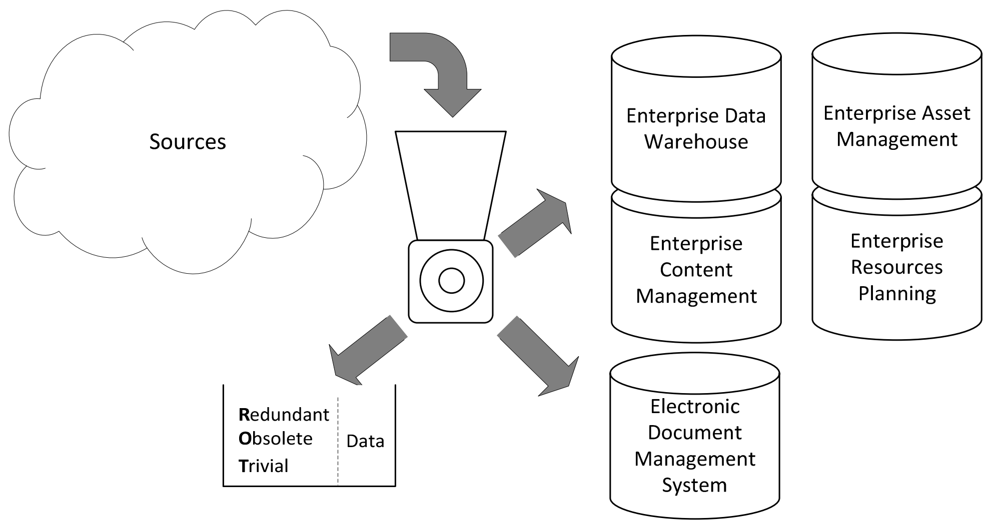


Figure 2.9: Data integration using proprietary applications

b) Manual method: Figure 2.10 illustrates an example of manual approach to data integration. The example shows a situation where an inventory is being developed from multiple heterogenous data sources. From the CAD model and material take off (MTO) spreadsheets, geometric attributes such as cut length and steel section specifications are extracted. From the maintenance database, information about repairs and replacements are extracted. Where a 3D laser scan model of the asset exists, scan positions of interest are also extracted. To carry out data extraction, these sources must be searched and navigated. Thereafter the values extracted are qualified and populated in another spreadsheet or database before any queries are applied for data extraction or analytics. In a scenario where there are hundreds of similar items e.g. steel sections, the assessor will need to go through the sources for each item - a process that is inefficient and error-prone, and likely to frustrate any benefits that may result from the task.

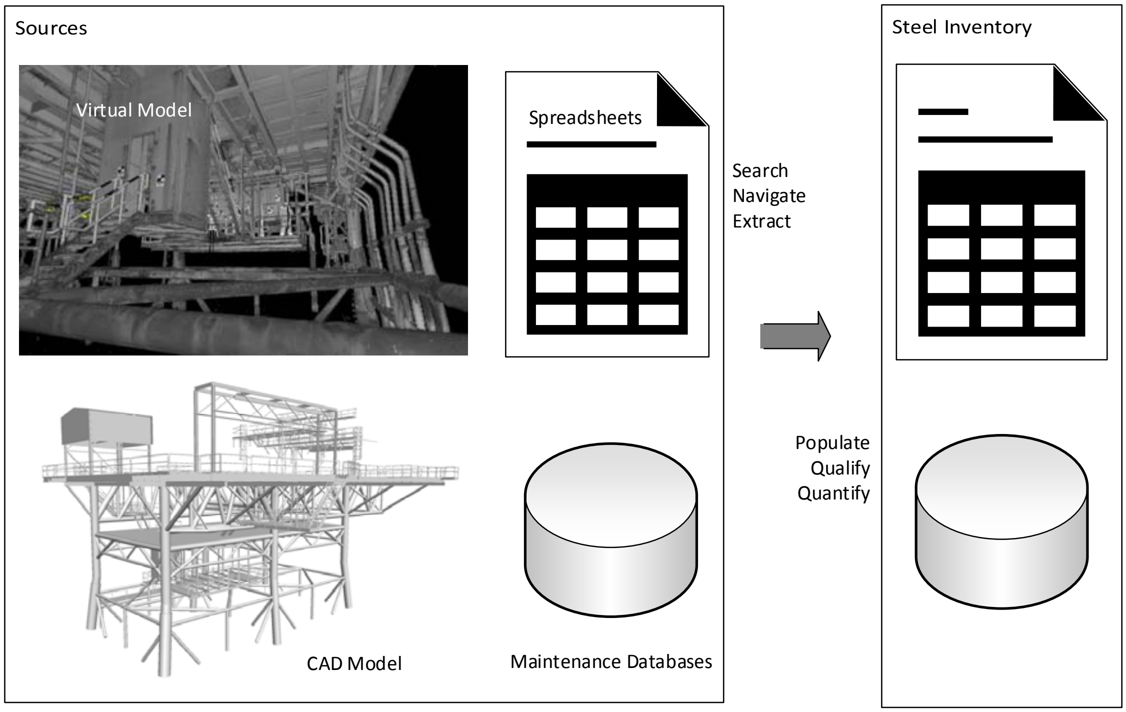


Figure 2.10: Example of manual approach for managing disparate data on projects

2.2.5 Data Integration Standard

a) Background: around 1980, several initiatives were taken by CAD users in the manufacturing industry to develop standards for the electronic exchange of technical data (Fiatech, 2011). To avoid a situation where CAD vendors may be confronted with too many incompatible standards, several national standardisation bodies decided in 1984 to develop a single international standard. This was taken up under the umbrella of ISO/TC184/SC4[1] and resulted ten years later in the publication of the initial release of ISO 10303[2], also known as STEP (Standard for the Exchange of Product Model Data) (Gielingh, 2008). During the years that followed the publication of ISO 10303, several more parts of STEP were published, aiming at usage in different industries. Of the standards developed for the exchange or sharing of product data, ISO 16739[3] for the AEC/FM industry and

[1] International Organisation for Standardisation Technical Committee 184 Subcommittee 4
[2] ISO 10303 *Generic Title*: Automation systems and integration — Product data representation and exchange
[3] ISO 16739 *Generic Title*: Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries

ISO 15926[4] for process industry are the most relevant for the built environment (Golzarpoor *et al.*, 2018). ISO 16739 is managed by ISO/TC59/SC13 in charge Organisation and digitisation of information about buildings and civil engineering works, including building information modelling (BIM), while ISO 15926 is being managed by ISO/TC184/SC4 in charge of industrial data.

b) ISO 15926: is detailed in Fiatech (2011) as the standard for integration of life-cycle data for process plants including oil and gas production facilities. It addresses the data integration challenges that are the focus of this research. However, the implementation of the standard is difficult because of the effort and skills required in using it (Wiesner, Morbach and Marquardt, 2011; Fiatech, PCA and USPI, 2015). Currently it is made up of twelve parts with varying levels of development. The standard is very important for oil and gas data integration because it provides an ontology - a shared conceptualisation - for the process industry domain using its Parts 2, 3 and 4. Part 2 documents the generic data model for representing technical information about process plant projects; Part 3 documents reference data for 2D and 3D geometric properties; and Part 4 documents the initial set of reference data for use with the Part 2 data model.

   ISO 15926 Parts 7 and 8 provide specifications for data exchange and life-cycle data integration. Part 7 specifies these using ISO 15926 templates based on Part 2 data model; and Part 8 does the same using Semantic Web technologies. Fiatech et al. (2015) describe ISO 15926 templates as standard addressable set of specific relationships between known things that together represent some information. Other parts of the standard include: Part 1 which introduces the standard; Parts 9, 10, 12 and 13 which are under development; and Part 11 which demonstrates the use of the published parts (Denno and Palmer, 2013). Table 2.1 describes the different parts of ISO 15926 and their statuses.

[4] ISO 15926 *Generic Title*: Industrial automation systems and integration -- Integration of life-cycle data for process plants including oil and gas production facilities

Table 2.1: Parts of ISO 15926

Part	Title	Description	Status
1	Overview and fundamental principles	States the purpose of 15926 - data integration using a data model that defines the meaning of information using common terms of the domain	International Standard (IS) ed.1 published June 2004
2	Data Model	Generic conceptual data model for representing technical information about project objects.	IS ed.1 published Dec. 2003
3	Reference Data for Geometry and Topology	Reference data for geometry; and topology for 2D and 3D shapes. It has ontology of basic classes of curves and surfaces that can be used with CAD and GIS and earth models.	Technical Specification (TS) ed. 1 published April 2009
4	Initial Reference Data	Initial set of reference data for use with the ISO 15926	TS ed. 1, published Oct. 2007
5	Procedures for Registration and Maintenance of Reference Data (withdrawn)	Procedures for registration and maintenance of reference data	Function has been taken over by a technical subcommitte for class library maintenance
6	Methodology for the development and validation of reference data	Describes how to validate a reference data item to ensure that it is genuine	Under Development
7	Implementation Methods for the Integration of Distributed Systems: Template Methodology	Specification for data exchange and life-cycle information integration using templates based on the data model of ISO 15926-2	TS published October 2011
8	Implementation Methods for the Integration of Distributed Systems: Web Ontology Language (OWL) Implementation	Specification for data exchange and life-cycle information integration using Resource Description Framework (RDF) and Web Ontology Language (OWL)	TS published October 2011
9	Implementation Methods for the Integration of Distributed Systems: Facade Implementation	Creation of private triple stores with ISO 15926 linked data	Under development
10	Conformance testing	Methods for defining and assessing conformance to ISO 15926	Under Development
11	Simplified Industrial Usage of Reference Data	Enables a flexible creation of product knowledge models that can be exchanged by combining RDF triples within named graphs, reference data dictionaries and a standardized set of relationships.	TS published May 2015
12	Life cycle integration ontology	Life cycle integration ontology in Web Ontology Language	Under Development
13	Integrated lifecycle asset planning	Integrated lifecycle asset planning	Under Development

c) ISO 15926 versus ISO 16739: considering the built environment covers all aspects of the environment built by humans, then process industry infrastructure should be a part of it. By implication ISO 15926 should be a subset of ISO 16736 since it is for the built environment. However, the two standards were created independently and there is no significant alignment between them (Gielingh, 2008). Their obvious similarities relate to the use of ISO 10303 Parts 21[5] (ISO,

[5] ISO 10303 Parts 21 *Specific Title*: Implementation methods: Clear text encoding of the exchange structure

2016), 28[6] (ISO, 2007a) and 42[7] (ISO, 2018a) in their implementation. ISO 10303 Part 21 allows information transfer using EXPRESS schema – based on standard EXPRESS modelling language in ISO 10303 Part 11[8](ISO, 2004). ISO 10303 Part 28 allows the transfer of geometric information and ISO 10303 Part 42 allows information transfer using conventional XML files conforming to an approved XSD Schema (Laud, 2013). ISO 15926 evolved from the European initiative EPISTLE[9] (Fiatech, 2011) while ISO 16739 was developed under the organisation International Alliance for Interoperability (IAI) – now buildingSMART (Abanda, Tah and Keivani, 2013). The latest version of ISO 16739 is the version 4 - the reason for the acronym IFC4.

Interoperability using ISO 15926 is based on STEP Application Protocol (AP) interoperability framework shown in Figure 2.11A and the corresponding parts of the standard are represented in Figure 2.11B. The key components of the framework include:

i. Application Activity Model (AAM), which describes the intended context for the exchange of product data. This considers the scope, information requirements and usage scenarios.
ii. Application Interpreted Model (AIM), an information model that specifies the normative part of the standard. AIM is a specialised subset of the Generic Resources: a set of information models containing constructs that can be shared by multiple APs.
iii. Application Reference Model (ARM), which describes the application view of the product data. The ARM is the mediator between the AAM and the AIM. Originally its purpose was to document high level application objects and the basic relations between them. But currently, ARM objects, their attributes and relations are mapped to AIM so that it is possible to implement an AP.

[6] ISO 10303 Parts 28 *Specific Title*: Implementation methods: XML representations of EXPRESS schemas and data, using XML schemas
[7] ISO 10303 Parts 42 *Specific Title*: Integrated generic resource: Geometric and topological representation
[8] ISO 10303 Parts 11 *Specific Title*: Description methods: The EXPRESS language reference manual
[9] European Process Industries STEP Technical Liaison Executive

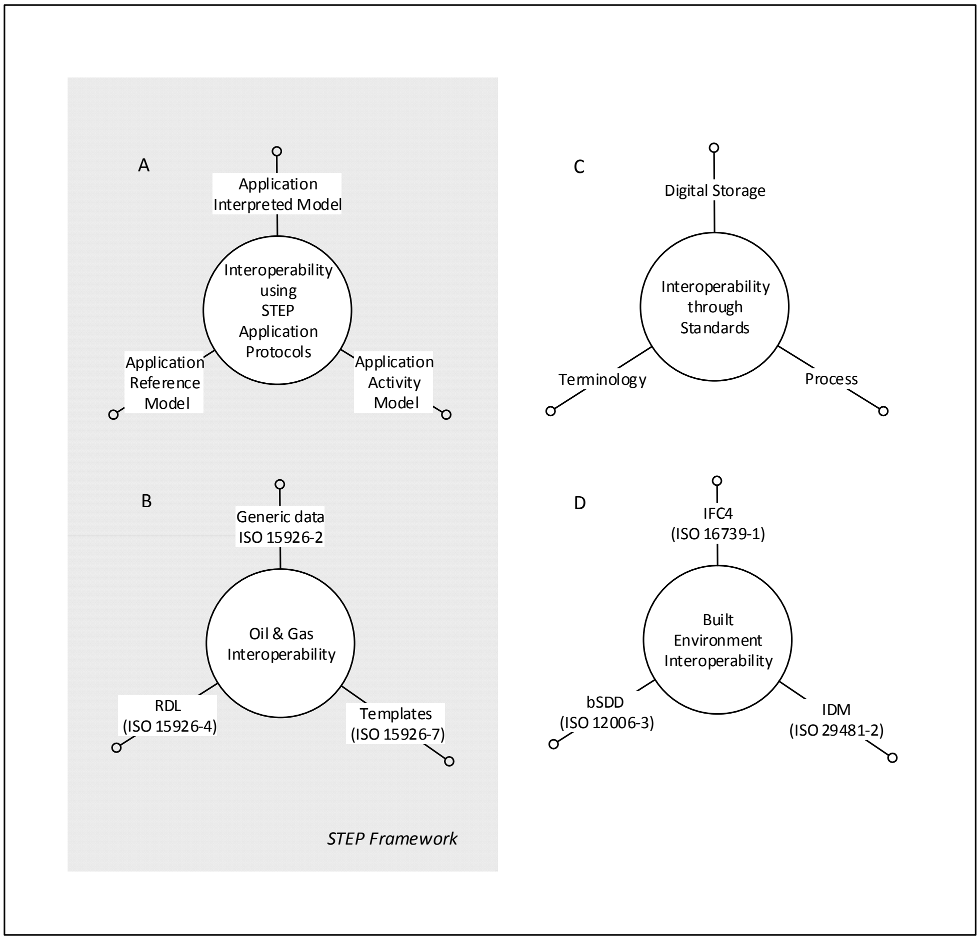


Figure 2.11: Interoperability through standards

In simplistic terms the three AP elements translate to *digital storage*, *terminology* and *process* represented in Figure 2.11C. These are essential for open interoperability standards. The goal of this structure is to avoid duplication of work and enable interoperability between APs of different industries and life cycle stages. This concept is replicated with different standards for the built environment sector as shown in Figure 2.11D. To achieve interoperability, ISO 12006-3[10] (ISO, 2007b) is used for mapping of terms. The standard defines the *terminology standard* called buildingSMART Data Dictionary (bSDD)[11] used for linking construction databases to BIMs (Volk, Stengel and Schultmann, 2014). ISO

[10]ISO 12006-3 *Title*: Building construction -- Organization of information about construction works -- Part 3: Framework for object-oriented information
[11] previously International Framework for Dictionaries (IFD) standard

16739-1[12] (IFC4) (ISO, 2018b) is used for data exchange and ISO 29481-2[13] (ISO, 2012) takes care of work process related information. The latter delivers Information Delivery Manuals (IDMs) which define scenarios and objects for exchange purposes.

Another difference between the standards relate to information transfer. ISO 15926 is designed for information sharing while ISO 16739 interoperability is by file exchange. However, the development of Semantic Web has improved the application of the two standards and both can now be use for integration, interoperability, and life-cycle data exchange (Abanda, Tah and Keivani, 2013; Golzarpoor *et al.*, 2018). To achieve this, ISO 15926-8 defines a format for the use of RDF/OWL XML to represent ISO 15926 schema. On the other hand ifcOWL (Abanda, Tah and Keivani, 2013) provides an OWL representation of ISO 16739 schema. Interoperability between the models of the two standards is also possible and has the potential to benefit:

i. suppliers whose products are used in both industries e.g. steel suppliers
ii. designers interested in the interactions between buildings and plants
iii. fabricators that deal with design information from both industries

However, this will depend on scope and level of detail for each industry as the absence of direct references in the meaning of the information between them will increase the risk of semantic heterogeneity and make interoperability harder. Available solution to this challenge requires mapping between the entities of the two standards e.g. XMpLant software (Nextspace, 2018).

2.3 Decision Analysis

This is the knowledge area that covers the theory, methods and practice of formal decision making. It also includes the tools and procedures required for identifying, representing and assessing a decision to give a decision maker the required insights into complex and uncertain situations (Covaliu, 2001). Figure 2.12 shows the relationship of the key

[12] ISO 16739-1 *Title*: Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries -- Part 1: Data schema
[13] ISO 29481-2 *Title*: Building information models -- Information delivery manual -- Part 2: Interaction framework

concepts of decision analysis relevant to this study. These concepts are described as follows:

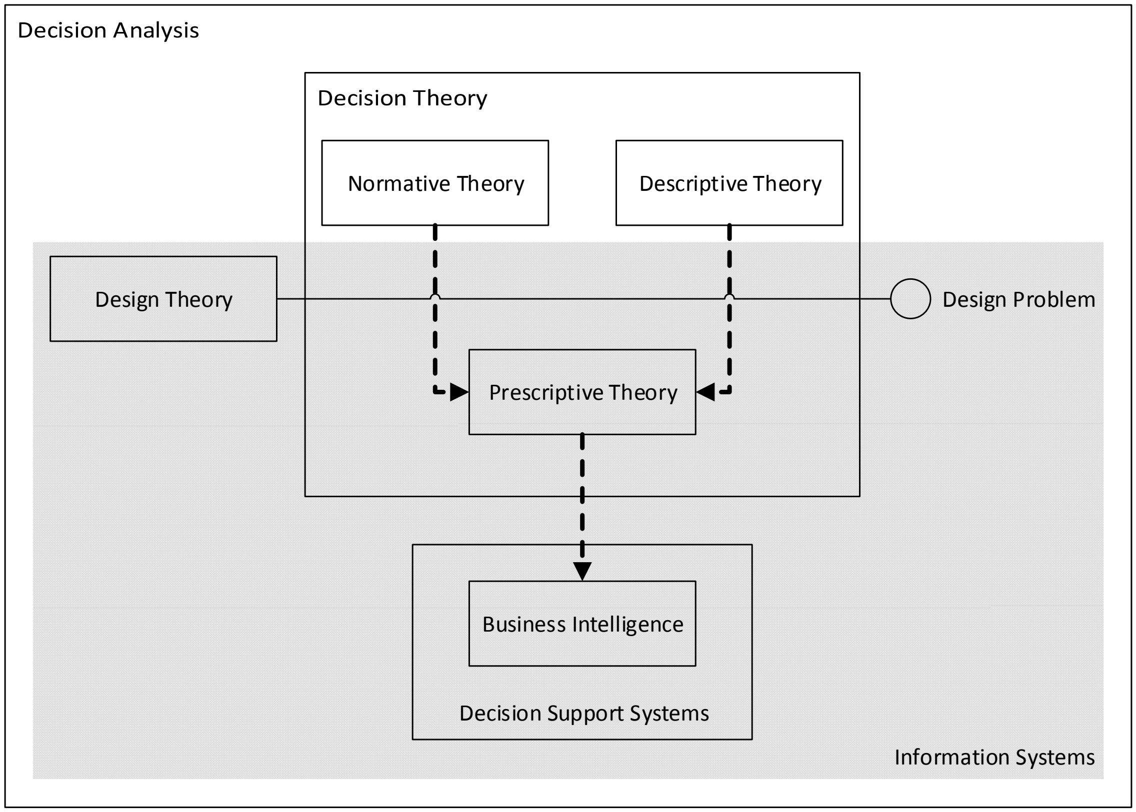


Figure 2.12: Anatomy of Information Systems Design Theory

a) Information Systems (IS): This is the discipline that is concerned with organisational interventions that help in the handling of data for organisational tasks (Hevner *et al.*, 2004). Examples of IS classes include DSS, executive information systems (EIS) and emergent knowledge process (EKP) (Al-Mamary, Shamsuddin and Aziati, 2014).

b) Decision Support Systems: This is a class of IS that supports decision making activities of organisations (Watson, 2009). It can be implemented at different levels of an organization - from regular operations to the executive management - and can be communications, data, documents, knowledge or model driven (Power, Sharda and Burstein, 2015).

c) Business Intelligence (BI): Watson (2009) defines BI as a DSS discipline concerned with the applications, technologies and processes for gathering, storing, retrieving, and analysing data in a timely and accurate manner in support of business decisions. BI is a theory performing domain of decision analysis in which normative theories of decision making are transformed into prescriptive theories through a thorough intellectual dialogue between academic and industry stakeholders (Clark, 2010). The theoretical focus of academic researchers, and the business objectives and limitations of practitioners inform the decision engineer's theory-backed practice. BI practice has multiple actors, viewpoints, theories and meta-designs for interventions (Watson, 2009). Clark (2010) argued that a good way to organise the practices within BI is to use the common distinctions observed by practitioners. Using this concept BI can be grouped as strategic, tactical and operational BI. Strategic BI supports executive level decisions, tactical BI supports management level decisions and operational BI supports process level decisions (IIBA, 2009). Figure 2.13 captures the key elements of a business intelligence system and they are described as follows:

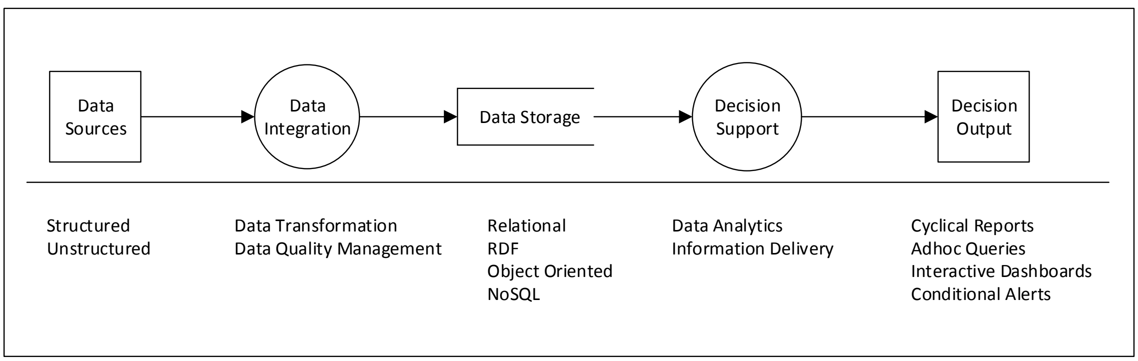


Figure 2.13: Business Intelligence System

i. Data Sources: data required for business decisions usually already exist within organisations and may or may not be in forms that are usable. Also, there may be challenges with retrieving them because of where they are stored (e.g. multiple locations), or how they are stored (e.g. inconsistent data dictionaries). Data in organisations can be broadly classified as structured, semi-structured and unstructured (Kitchin, 2014). Structured data are common in traditional organisations and usually make use of approved data dictionaries and rules-driven templates to ensure information integrity. Discovery using this type of data is supported by well-defined data models and business rules. Examples of structured data include relational databases and data stored in spreadsheets. That said, some processes within organisations do not comply with the way structured data are generated and managed. As a result, data with varying levels of structure are also generated. Examples include XML which is semi-structured, and text, images, audio and video, which are unstructured. For unstructured data, the structure and relationships of the entities are not predefined, and no specific organisation rules have been applied to ensure data integrity. Also, discovery from this type of data involves defining metadata and developing data matching algorithms to extract information (Rusu *et al.*, 2013).

ii. *Data Integration:* is the service described in Section 2.1. It is required for connecting the sources to the target systems. Important considerations for data integration include how and where (technologies) data will be transformed as well as when (frequency and latency) data will be extracted. Data integration should also comply with applicable business rules to ensure decision outcomes that are representative of user's process. Lastly, data integration is dependent on the quality of the data in the sources. The reliability of such data depends on their completeness and validity. As a result, additional verification to correct error values and complete missing ones is essential (Doan, Halevy and Ives, 2012a).

iii. *Data Storage:* is concerned with where data is physically stored. In BI systems, storage is provided by a Database Management System (DBMS) configured to the specific needs of a business environment. Established data storage technologies applicable in BI systems include relational database (RDB) (Connolly and Begg, 2005), Resource Description Framework (RDF) (Cyganiak, Wood and Lanthaler, 2014), object databases (Doan, Halevy and Ives, 2012b) and NoSQL databases (Kitchin, 2014). RDB stores data in tables and uses Structured Query Language (SQL) as the standard query language. RDF is a graph-based model with globally defined identifiers and terminologies. It is queried using SPARQL (The W3C SPARQL Working Group, 2013). Object databases store data as objects and are particularly useful for storing spatial data. NoSQL databases provide the flexibility absent in relational databases e.g. scaling out, fluid schema and nonstandard query language (Connolly and Begg, 2005).

iv. *Decision Support:* decision support information can be applied at multiple levels or in specific areas of an organisation (IIBA, 2015). What is important is that the data that form decision information are coordinated appropriately from the different sources and undergo correct analysis. BI systems usually support descriptive, predictive or prescriptive analytics (Ryu, 2013). Descriptive analytics makes use of historical data for understanding past outcomes. Its focus is on the information and communication requirements for standard and ad hoc reporting. For this type of analysis, there is no consideration of potential decisions, actions or situations. Predictive analytics makes use of statistical analysis methods to identify patterns in data and then uses that knowledge to forecast future events. Its focus is on the information requirements for pattern recognition and considers specified situations and business rules in making predictions. Prescriptive analytics extends predictive analytics. It does this by recommending appropriate actions to be taken after predictions. Its focus is on statistical optimization and simulation techniques, and considers everything (business rules, constraints, risks etc.) related to a decision. Visualisation techniques are useful for understanding all outputted decision support information. Consequently, presentation of analytical insights in BI systems is usually supported by reporting facilities that ensure effective display of decision outcomes.

v. *Decision output:* examples of how decision outputs are communicated include predefined and ad hoc reports, dashboards, balanced scorecards, queries using online analytical processing (OLAP) (Horakova *et al.*, 2013), and conditional alerts. Decision output can be set to be selected from specified data dimensions and levels of granularity. Such an output can also be further explored with the filtering functions of a BI system where they exist. Other BI system functions that can improve the value of decision output include translation, computation and data aggregation.

d) Decision Theory: According to Steel and Stefánsson (2016), decision theory is concerned with the cognitive activity behind the choices of an entity - usually an individual person capable of deliberation and action. These choices can be as simple as deciding on who to vote for or as complex as deciding to pursue a political career. The general premise is that whatever a decision maker decides on at any given time is influenced by beliefs, desires and preferred values. Clark (2010) categorises decision theory into three:

i. *Normative Theory*: This is the domain of mathematical economists and the focus is on how a supposed rational decision maker makes perfectly informed choices. In practice, the theory's concern is on 'how decisions should be made'.
ii. *Descriptive Theory*: This is the domain of organisation theorists, and the focus is on the observed behavior of a decision maker. In practice, the theory's concern is on 'how people actually make decisions'.
iii. *Prescriptive Theory*: This is the domain of the practitioner and the focus is on finding DSS to help decision makers arrive at better choices in line with normative theories, within their cognitive limitations. In practice, the theory's concern is on 'how to improve the quality of decisions'.

BI is an output of prescriptive theory. It bridges the gap between normative theories of how decisions should be made and the realities of how they are made. BI developers' challenge is how to create sociotechnical solutions that improve the quality of decisions (Cabantous and Gond, 2010). They approach this prescriptive challenge by designing recipes for developing specific IT solutions and theorising about the phenomena around the IT artefacts (Gregor, 2006). This process is captured in a model that inputs normative and descriptive theories, and outputs prescriptions. One view of such model is *design theory* used in IS domain (Clark, 2010).

e) IS Design Theory: IS design theory refers to a set of concepts, beliefs, conjectures and generalized scientific laws (both natural and social) by which systems developers effectively map design problems to solutions for a special class of IS (Hanseth and Lyytinen, 2004). A good design theory enables effective transfer of associated knowledge to new scenarios (Hevner *et al.*, 2004). According to Hanseth and Lyytinen (2004) the propositions of a design theory include:

i. a set of requirements for a family of design problems,
ii. a set of system features (or principles for selecting them) that meets the requirements,
iii. a set of principles useful for guiding the design process

These elements combine to generate a set of guidelines, through the lens of established theories, used for implementing new designs. Table 2.2 describes the focus of these elements.

Table 2.2: Design theory components (Hanseth and Lyytinen, 2004)

Elements	Focus
Requirements/ goals	Describes the class of goals to which the theory applies
A set of system features	A set of IT artifacts (class) hypothesized to meet the requirements.
Design Theory	Theories from natural and social sciences governing the design requirements or the processes arriving at them.
Design principles	A codification of procedures which when applied increase the likelihood of achieving a set of system features. These procedures are derived logically from kernel theories

g) *Kernel Theory*: Kernel theories are design theories sourced from natural and social sciences, and used as hermeneutics (method of interpretation) in generating meta-designs for classes of IS (Hanseth and Lyytinen, 2004; Clark, 2010). Decision engineers use Kernel theories to seek theoretical perspectives and the corresponding prescriptions in the development of new IS. There are multiple Kernel theories that support decision making rationality in real world scenarios. Nonetheless, Clark (2010) concludes three important Kernel theories are useful in grouping the variety of solutions in the BI domain. These theories include:

i. theory of bounded rationality for decision development: This is focused on selection from among alternatives, given limited information, time, and attention. It is typically used for decisions taken at the operational level.
ii. theory of distributed cognition for capability crafting: This is focused on deriving decisions from interdependent interactions of individuals in a group

context characterized by culture, tools, processes etc. It is typically used for decisions within a tactical endeavor.

iii. organised anarchies theories for issue illumination: This is focused on extracting decisions from a group that has members with differences in political views, abilities, interests etc. by making them reach a compromise. This is characteristic of strategy-making processes.

An important note is that the described kernel theories can possibly be applied at any level of an organisation, depending on the prevailing context. For this research, the theory of bounded rationality is applicable as the focus is on decision development.

h) *Bounded Rationality*: Simon (1972) explains bounded rationality as the concept that takes into account the limitations affecting the rationality of individuals making decision. These limitations include the ability to track the decision problem, cognitive limitations of the decision maker, time available for the decision to be made and computational limitations. The decision maker in this context acts in a satisficing manner and only seeks a solution that is satisfactory and not necessarily the optimal. The theory of bounded rationality is an upshot of the theory of perfect rationality (Miljkovic, 2005) and its focus is on how decision makers choose from alternatives. Simon (1972), in addition, states that it includes descriptive theories in areas including psychology and economics; and prescriptive traditions in operations research and decision support systems.

The theory of bounded rationality has three assumptions: the existence of a decision maker, a finite set of alternatives and a decision challenge (Clark, 2010). However, the focus is not on who the decision maker is, what the selected alternative is or when a decision is reached. Rather, the focus of the theory is on how the decision was made - which is contingent on the search technique. This makes search the core concept of bounded rationality theory. Decision makers are sometimes not aware of the range of alternatives and some information may as well be missing. Consequently, the search process for the selection of an alternative is an active one and guided by heuristics.

i) *Heuristics*: This is a set of rules on ways to search. Clark (2010) describes it as an 'adaptive tool kit' of computational strategies that allow the exploration of the information environment by the decision maker. Rules range from what to choose to how to search and when to stop. The application of heuristics helps decision makers to reach accurate decisions in a short time.

2.4 Summary

The first part of this chapter introduced the basic concepts of data integration, its main approaches, and the common solutions applied in the built environment sector. In addition, it discussed the Semantic Web and described a simple example on how it works. The second part of this chapter described process industry elements including: the process features, stakeholders, infrastructure, and life cycle stages. It also exposed the common data sources in the industry. Thereafter, it described some of the life cycle activities that can significantly benefit from data integration and discussed the data quality dimensions that should be addressed to gain the benefits. In addition, current process industry practices for data integration were explained, and the industry recognised data integration standard discussed. To conclude, the areas of decision analysis relevant to this study were discussed. The next chapter explores a use case in developing a business case for data integration in offshore decommissioning.

Chapter 3

Business Case

3.1 Offshore Decommissioning Process

The offshore decommissioning process as described by Arup et. al (2014) usually takes place in multiple phases and can last for years. The duration depends on the speed of permitting approval, availability of required marine vessel and equipment spreads, type and size of offshore structure, and availability of skilled personnel (Siems, 2016). Appendix 1 shows a detailed breakdown of the activities involved in offshore decommissioning process and they are summarised as follows based on groupings that reflect resources management.

a) Project Management and Monitoring: Related activities are managed by the asset owner or operator. Activities include the planning, engineering and scheduling of the tasks required for developing the decommissioning plan particularly for regulatory permitting. It also includes the permitting process and overall overseeing of the entire decommissioning process. In addition, it includes monitoring activities beyond the decommissioning process.

b) Well Abandonment: This is managed by a spread of specialist equipment and associated personnel. The activities involved include plugging of oil wells with cement and flushing of the pipelines – those connecting the oil wells and those connecting one platform to another – with seawater. For flushing, pigs – intelligent robotic devices for evaluating the interior of the pipe (Devold, 2013) are propelled through the pipelines to remove residual hydrocarbons, and the pipelines are filled with seawater corrosion inhibitors until removal or final desertion at sea bed.

c) Make-Safe Operations: This can be managed by a mobilization of specialist equipment and accompanying personnel. Activities include identification and correction of hazards that can be problematic on the facility, and general preparation of the asset for decommissioning operations. Examples of preparation activities include repairs to hand rails and deck gratings, preparation of habitable living facilities for workers, servicing of lifting cranes to be used, disconnection of the wells from the process equipment, cleaning and treating of production equipment and piping to rid them of hydrocarbons and other hazardous materials.

d) Subsea and Topside Removal: This is usually managed by a spread of specialist equipment and related personnel. The activities involved are to sever and remove the well conductors, platform topside, and platform jackets underwater. Also included are the activities to disconnect the topside components for removal. The removed items are shipped to shore for onward disposal in an environmentally responsible way. In some cases, the jackets may be deposited at an approved reef site (DECC, 2011).
e) Subsea Site Remediation: This is another vessel mobilization, usually a trawl vessel, to clear the site of the decommissioned facility and verify that there are no obstructions.
f) Topside and substructure recycling: Once the decommissioned facility gets to shore, it is handled by a dismantling contractor who will take it to pieces and send it for recycling or store it for future reuse.

3.2 North Sea Decommissioning

Over the next 30 years, more than 470 offshore installations in the North Sea's UK Continental Shelf will need to be decommissioned (Oil and Gas UK, 2014). According to the Royal Academy of Engineering (2013) determining the cost of this is extremely difficult because of the volatility in estimated risks, material condition and oil price. In addition to the cost issues, DNS et al. (2015) identified limited industry experience and inadequate human resources as challenges. While offshore decommissioning presents a significant business opportunity to UK companies, the owners of the assets and the UK Government are liable for the cost as the UK Government will be granting tax relief to the asset owners (HM Revenue & Customs, 2012). As a result, the UK tax payer is a very important stakeholder. It is therefore very important that the costs are kept as low as possible. The main drivers, issues and opportunities associated with offshore decommissioning in the UK North Sea are highlighted as follows.

a) Drivers

i. *End of Life:* According to the Royal Academy of Engineering (2013), oil and gas production in the UK Continental Shelf (UKCS) started in the 1970s. Considering that a significant number of the facilities have a design life of 25 years (Stacey, Sharp and Birkinshaw, 2008), then facilities installed prior to the 1990s may be due for decommissioning. Life extension programmes are currently employed to keep facilities in operation beyond their design lives (Executive and Health Saftey, 2014). However, with continued usage, Stacey et al. (2008) state that challenges with asset integrity increases and the decision to decommission becomes inevitable.

ii. *Return on Investment:* With the drop in crude oil price (Nasdaq, 2017), decommissioning is expected to pick up. This is because asset owners are unwilling to invest in life extension programmes for old assets, as they may not be able to recover the cost (Deloitte, 2015). In addition, when it is not viable to produce oil for reasons other than market price e.g. low oil yields, production assets are prioritized for decommissioning (Oil and Gas UK, 2015a).

b) Issues

i. *Asset Information*: As described in Yang (2009) and DNS et al. (2015), from the conception of an asset to the point where the decommissioning decision is made, various stakeholders capture significant amount of disparate data about the asset in various databases and documents. As shown in Figure 3.1, design, construction, operations, maintenance, revamp and as-built data are collected over the life cycle of an asset and are needed for decommissioning. The data assembled is required for materials management, energy estimation, cost analysis, schedule development etc. (DECC, 2011). To determine the condition of a decommissioned item for example, information on design, performance in operation, and maintenance records will be required in coming to a judgement about its residual strength or liabilities. However, at decommissioning, some of the required information may be missing and the available data is likely to be in disparate formats. This will make the assessment process very difficult. Also considering that there are large numbers of similar items on oil and gas facilities, the process is prone to error. The assessors will need to carefully string the available information together to make a determination on reuse or recycle.

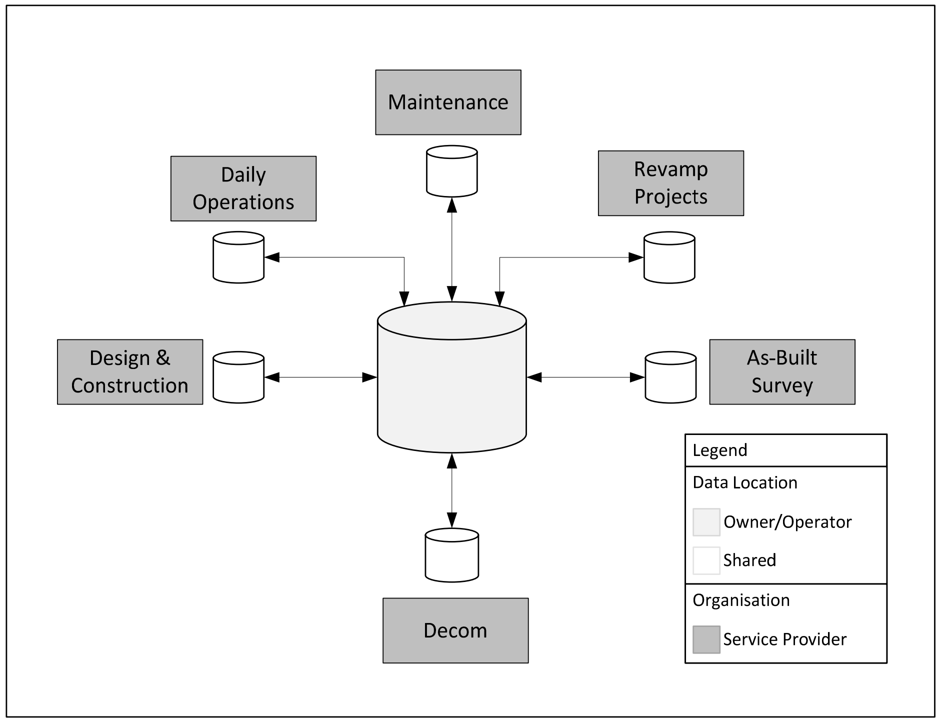


Figure 3.1:Typical data flow for offshore rigs

ii. *Safety*: Offshore decommissioning involves the removal of heavy structures from the world's most inhospitable environment - the sea. The process poses significant health and safety challenges considering that typical operations in that environment are associated with hazards of fatal consequences (Christou and Konstantinidou, 2012). As a result, decommissioning programmes require strict health and safety considerations (HSE Energy Division, 2014) and full compliance with applicable safety regulations (Bureau Veritas, 2011). Estimates from previous decommissioning projects, made on the statistical probability of serious and fatal accidents occurring during the decommissioning process, e.g. the 'Ekofisk I' (13 jackets and a tank), indicated that the Potential Loss of Life (PLL) during decommissioning can be up to 8% (Ekins, Vanner and Firebrace, 2005). This is high, usually of concern, and the basis why offshore personnel mobilization and lodging are typically recommended to be as low as practicably possible.

iii. *Regulations*: The key consideration of this research with regard to regulations is stated in the guidance by DECC (2011) and it relates to the requirement that the topsides of all offshore oil and gas installations must be returned to shore for reuse, recycling, or final disposal on land. This is to ensure they are managed appropriately in an environmentally responsible way. Complying with regulations where facilities have significantly deteriorated structurally may require single lift removal by special heavy-lift vessels. As these vessels are few, hiring costs are high and engaging them will require long lead times (Dalgic, Lazakis and Turan, 2013). These waiting periods can impact the decommissioning schedule and increase the preservation cost of the facilities where production has ceased.

c) Opportunities

In order to sustain the oil and gas industry in the UK North Sea, the UK Government is strategically supporting the development of the decommissioning industry by advising on (i) development of late-life business models and (ii) ways to eliminate the barriers to cost effective decommissioning (Oil and Gas UK, 2015b). This strategic support, in part, is to leverage the job creation opportunities inherent in decommissioning in order to recover some of the jobs that may have been lost in the exploration and production business areas due to investment challenges. Estimates of decommissioning jobs across the North Sea till 2040 is valued at£46bn (RSA and ZWS, 2015) and the main business areas can be grouped into (i) services for removal of a facility from its offshore location and (ii) waste management on land (ARUP, Scottish Enterprise and DNS, 2014). With regard to the removal process, study by DNS et al. (2015) indicate that several of the existing service providers can adapt their tools and techniques to effectively support the required functions. However, RSA and ZWS (2015) and DNS et al.(2016) indicate that waste management of decommissioned offshore facilities under prevailing regulations requires innovative ways to exploit the inherent benefits.

Chapter 3: Business case

According to DNS et al. (2015), current waste management initiatives in decommissioning programmes employ the circular economy concept because it provides an avenue for cost reduction. The objective of a circular economy is to keep a product in use for as long as possible and extract maximum value from it in the process (Ellen MacArthur Foundation, 2015). This allows items from a decommissioned facility to be used in other systems and create value by either saving on cost, when applied directly by an owner, or earning cash when sold. In the UK offshore decommissioning industry, the following waste management techniques are often in consideration:

i. *Material recycling*: This process involves removing the materials and components of the decommissioned assets and processing them to change their forms for use in a new entity. Crushing concrete to aggregate and smelting of steel materials are good examples. These conversion processes are carbon intensive, so not a favorite with respect to environmental consideration. Currently, the absence of local capacity for steel smelting in the UK highlighted in RSA and ZWS (2015) and DNS et al. (2016) means the value created from the process is lost to other countries. The opportunity that exists here is the need for local capacity for steel recycling in the UK.

ii. *Component re-use*: Cleaned Items from a decommissioned facility can be reused directly in similar or totally different functions without having to rework them at all. It could be that the items are brand new spares, have been maintained properly or have just been recently replaced. Examples of items that can be reused include structural steel sections - for piling and as railway sleepers (DNS, AMEC and ZWS, 2016), electric motors and valves (RSA and ZWS, 2015). DNS et al. (2015) also indicate the possibility of reusing complete modules e.g. complete helideck and test separator systems.

iii. *Repair and re-use*: After assessment of equipment and components from a decommissioned facility, some of them will require some repair, sub component replacement or sometimes recertification before they can be reused. Examples identified in RSA and ZWS (2015) include winches, steel sections and water treatment filtration units.

Considering the large percentages of recycled decommissioned materials in DNS et al. (2015) and DNS et al. (2016), it can be concluded that asset owners seeking to meet their environmental obligations often adopt this method. However, estimates indicated in RSA and ZWS (2015) show that increasing re-use over recycling can increase the value of items by between five and seven times, which roughly adds up to a third of the total value of an installation at the decommissioning stage. According to the guidance by DECC (2011), reuse is also a good strategy in gaining government approval for decommissioning programmes and can help to save on overall decommissioning cost. Consequently, this research focuses on tools to achieve reuse of waste from decommissioned facilities.

3.3 State-of- the-art tool for decommissioning waste management

The Late Life Planning Portal (DNS, 2017) is a repository for knowledge sharing on decommissioning projects in the UK North Sea oil and gas industry. The platform documents best practices and tools for decommissioning practitioners. The recommended tool for decommissioning waste management is a tool by D3 Consulting Limited, a specialist decommissioning waste management organisation (D3 Consulting Limited, 2017). The tool is a spreadsheet template with attributes associated with waste items registered as column headers. To use the tool, all the attributes related to an item have to be observed from the relevant sources and logged in the correct column cells on the corresponding rows of the item. After the tool has been populated with the required data, spreadsheet functions are then used to manipulate the data to sort, count etc. Figure 3.2 shows a snapshot of the tool. This approach requires vast amounts of costly error-prone human effort.

Asset	Topside/ Jacket/ Subsea	Functional Category	Description	EWC	EWC Description	EWC Classification	Net (Te)	Units	Source

Figure 3.2: Spreadsheet template for offshore decommissioning waste management

3.4 Knowledge Gap

Over the years, data integration technology applications in the oil and gas industry have focused on meeting the needs of asset life cycle stages other than decommissioning

(Fiatech, 2011). This is the case because most assets are just reaching the end of their design lives (Royal Academy of Engineering, 2013). Currently, limited work has been done on applying the concept of data integration to legacy oil and gas data for offshore decommissioning waste management purpose (D3 Consulting Limited, 2017); and a report by industry stakeholders underscores this need (DNS, AMEC and ZWS, 2016). A review of the research literature (Table 3.1) shows that ontology-based methods can effectively handle the challenges associated with process plant data integration but are yet to be applied in the decommissioning process. Proximally situated to the latter, is the evidence of the application of other technologies (e.g. the World Wide Web and relational database) for waste management in the building industry (Table 3.2). However, none of the studies demonstrate use cases and implementation plans that can be easily adapted. Considering that the currently recommended method for decommissioning waste management in the UK offshore decommissioning industry has a manual approach, this research applies the well-established ontology-based concept for data integration in offshore decommissioning waste management process.

Table 3.1: Related research based on data integration technology

Source	Summary	Data Integration Resource	Industry	Life Cycle Stage	Integration Technology: Ontology	Integration Technology: Semantic Web	Integration Technology: Others	Industry: Oil & Gas	Industry: Other Process Industries	Industry: Others	Lifecycle Stage: Engineering	Lifecycle Stage: Procurement	Lifecycle Stage: Construction	Lifecycle Stage: Operations	Lifecycle Stage: Maintenance	use case: Decommissioning	use case: Waste Management
(Rong, 2009)	This research demonstrated the use of ISO 15926 and cloud computing for information integration in large scale oil and gas industry operations that involve multiple disciplines and inter-enterprise collaboration.	ISO 15925 Cloud Computing	Oil and Gas	Operations	✓		✓	✓						✓			
(Bravo et al., 2011)	This research proposes and demonstrates an information system architecture that applies, amongst other technologies, an ontological framework that enables semantic integration to ensure effective information exchange between applications used in oil production process.	Ontology Semantic Web	Oil and Gas	Operations	✓	✓		✓						✓			
(Kim et al., 2011)	This research presented a prototype ontology-based repository that makes use of ISO 15926 and Semantic Web technologies for storing and sharing equipment data with repositories within its confederation.	15926 Semantic Web	Nuclear Plant	Operations	✓	✓			✓					✓			
(Wiesner et al., 2011)	This study presented an ontology-based approach for information integration in chemical process engineering. The approach, implemented as a tool, makes use of a knowledge base that reconciles the ontologies of the data sources, and semantic technology for interrogating the knowledge base.	Ontologies Semantic technology	Chemical Plant	design	✓	✓			✓		✓						
(Lee et al., 2012)	This study analyses various data models for integrating lifecycle data in different domains and develops information requirements for combined management of facility, process and output data of a manufacturing system. Using the developed information requirements, an appropriate data model was generated and used to demonstrate how to effectively manage the operations and maintenance activities of manufacturing systems.	15926 10303	Manufacturing	O&M	✓				✓					✓	✓		
(Fiorentini et al., 2013)	This study reports on the use of ISO 15926 for interoperability in a Product Life Cycle Management project involving multiple information systems between power industry organisations.	ISO 15925	Nuclear Plant		✓				✓								
			Research Focus		•	•		•								•	•

Table 3.1: Related research based on data integration technology *(Continued)*

					Integration Technology			Industry			Lifecycle Stage						use case
Source	Summary	Data Integration Resource	Industry	Life Cycle Stage	Ontology	Semantic Web	Others	Oil & Gas	Other Process Industries	Others	Engineering	Procurement	Construction	Operations	Maintenance	Decommissioning	Waste Management
(Borgo, 2014)	This study investigated the challenges associated with expanding standard ontologies in industrial domain and proposes an approach to deal with new dependency relations without affecting the information system's overall consistency.	15926 DOLCE	Industrial System		✓			✓	✓								
(Novák et al., 2014)	This study proposes and demonstrates a platform for the integration of simulations and complex data access systems. The integration platform makes use of ontologies and semantic web technologies for capturing engineering knowledge leveraged for the simulation models.	Ontologies Semantic Web	Water distribution	Design Operations	✓	✓			✓		✓			✓			
(Novák et al., 2015)	This study uses Semantic Web technology to develop an ontology for integrating heterogenous data sources and service-oriented tools as knowledge base for the support of dynamic simulation in industrial plants.	Ontologies Semantic Web	Water distribution	Design Operations	✓	✓			✓		✓			✓			
(Ebrahimipour and Yacout, 2015)	This study described a dynamic and interactive ontology-based platform that makes use of ISO 15926 and Semantic Web technologies for the integration of data from multiple sources. The developed knowledge base is used to support decisions that ensure optimised use of equipment across the life time of oil and gas plants.	15926 14224	Oil and Gas	Operations	✓			✓						✓			
(Šindelář and Novák, 2016)	This research describes an ontology-based approach for integrating the simulations and models of an industrial process in an automation system. The approach makes use of several domain ontologies, including ISO 15926, for its knowledge base. It also uses Semantic Web technologies for extracting information and detecting inconsistencies in the knowledge base.	Ontologies Semantic Web	Industrial System	Design Operations	✓	✓		✓	✓		✓			✓			
			Research Focus		•	•		•								•	•

Table 3.2: Related research based on waste management use case

Source	Summary	Data Integration Resource	Industry	Life Cycle Stage	Integration Technology: Ontology	Integration Technology: Semanmtic Web	Integration Technology: Others	Industry: Oil & Gas	Industry: Other Process Industries	Industry: Others	Lifecycle Stage: Engineering	Lifecycle Stage: Procurement	Lifecycle Stage: Construction	Lifecycle Stage: Operations	Lifecycle Stage: Maintenance	Lifecycle Stage: Decommissioning	use case: Waste Management
(McGrath, 2001)	This resource is a proprietary tool that enables pre-demolition audit. The tool is a software application targeted toward the construction industry but which can be applied in other domains. To use the tool, waste data needs to be pre-prepared in a recommended template and then imported into the software. Otherwise, attributes of each waste element will need to be entered manually using a form within the software	Proprietary tool	Building General	Construction Demolition			✔			✔			✔			✔	✔
(Banias et al., 2011)	This study presented a construction and demolition waste management decision support system that makes use of a relational database management system for storing waste data; the World Wide Web technologies for delivering web pages to user requests; an algorithmical model for the estimation of generated wastes; and mathematical programming for specifying optimal waste management strategies.	WWW RDB	Building	Construction Demolition			✔			✔			✔			✔	✔
(Li and Zhang, 2013)	This study proposed a construction waste management system that makes use of a web browser for user interfacing; an application logic for integrating the accounts, data input and analytical modules; a web server for delivering internet web pages; and a relational database management system for storing waste data.	WWW RDB	Building	Construction			✔			✔			✔				✔
(D3 Consulting Limited, 2017)	This resource is a freely available tool contributed by an offshore decommissioning waste management service provider for decommissioning waste audit. The tool is a basic spreadsheet that needs to be populated with waste data before use. The tool is a spreadsheet and has no special functions programmed into it.	Manual	Oil and Gas	Decomissioning			✔	✔								✔	✔
			Research Focus		•	•		•								•	•

3.5 Summary

This chapter explained the offshore decommissioning process. Thereafter it described the drivers, issues and opportunities of offshore decommissioning in the UK North Sea. It also described a gap analysis carried out to investigate the application of ontology-based data integration for offshore decommissioning waste management. From the analysis, it showed that the current practice for data integration is a manual process. It discussed this solution and argued for the application of ontology-based approach considering the evidence that it can effectively handle the challenges associated with built asset data integration. The next chapter discusses the theory and methods used for this study.

Chapter 4

Methodology

The aim of this research is to investigate the current decision-making environment in offshore decommissioning waste management and find ways of improving existing practices with the application of information technology. To achieve this, the following objectives must be achieved:

i. analysis of the offshore decommissioning process
ii. analysis of existing data integration approaches
iii. design of a data integration system with querying capability
iv. development of a life cycle data integration framework
v. validation of developed data integration framework with prototypes

The first objective - documented in Chapter 3 - helps in identifying a knowledge gap and the remaining four help in developing the appropriate solution. This research pattern is analogous to abductive research approach (Bryman and Bell, 2015). This research approach addresses the shortcomings of deductive and inductive research approaches (Riemer, Lapan and Quartaroli, 2012) using the pragmatist philosophical stance (Onwuegbuzie, Johnson and Collins, 2009) which seeks to solve research problems in a practical way. Deductive research approach is not well suited because this study lacks a well-defined theory that can be tested. Inductive research approach on the other hand requires using enough empirical data to formulate a theory. However, the domain concerned in this study is so large that it is only practicable to address use cases. Abductive research approach addresses these weaknesses by assuming that with some research data it is possible to make logical inferences and construct theories. Its research process starts with identifying a fact of interest and concludes by explaining it. As a result, it is adopted for this study.

4.1 Theory for this research

The perspective used for this research is that of Business intelligence (BI). Based on the model described in Table 2.2 and the IS design theories related to BI, an inspection of the aim of this research indicates the need to support a decision process. As a result, bounded rationality theory applies. This means the typical features of bounded rationality theory also apply. A quick review of the objectives of this research highlights the need for data integration and query answering – these are the focus of the principles for design. Table 4.1 captures the design theory for this research using the model described in Table 2.2. It is important to note that the cognitive limitations of the decision maker are not part of this study.

Table 4.1: Design theory components for this research

Elements	Focus
Requirements/ goals	Decision development
A set of system features	Decision problem One decision maker Cognitive limitations of the decision maker Limited decision time
Design Theory	Bounded rationality Heuristics
Design principles	Data integration and search

4.2 Research Process Model

This research is divided into four buckets of activities as shown in Figure 4.1. The first set of activities aligns with the first research objective and is for the development of a business case for data integration in offshore decommissioning. To develop the business case, processes in the oil and gas domain with data integration challenges are identified (Step 1). The challenges with these processes are used to identify the needs (Step 2). A domain use case that captures the needs identified is then used to elicit solution requirements from stakeholders (Step 3). This process considers the desired outcomes, assumptions, risks and constraints.

The second bucket of activities aligns with the second research objective and focuses on the analysis of the requirements for the desired solution. After the set of requirements elicited from the stakeholders have been correctly validated and verified, solution approaches that address the identified needs are researched and reviewed (Step 4). This process helps to identify the most appropriate approaches to address the needs. The most fitting options of the identified approaches are then matched to the stakeholders' requirements (Step 5). The resulting analysis is used to conceptualise a solution design to address the need of the business case (Step 6).

The third bucket of activities aligns with the third and fourth (partly) research objectives and is focused on the development of a working but limited version of a solution design. To achieve this, a thorough study of the technical resources (technologies and processes) required to construct solution artifacts is carried out (Step 7). Thereafter, the resources are configured and tested out in a prototype of the solution design using a simple use case (Steps 8 and 9).

The last bucket of activities aligns with the fourth (partly) and last research objectives and is focused on the evaluation of the design prototype using multiple use cases (Steps 10 and 11). The solution evaluation tasks are designed to assess, through measurement and analysis, the performance of and value delivered by the solution design when applied in use cases of interest and compared to corresponding existing practices (Step 12). They are also used to identify limitations of the proposed solution and recommend approaches for improvement.

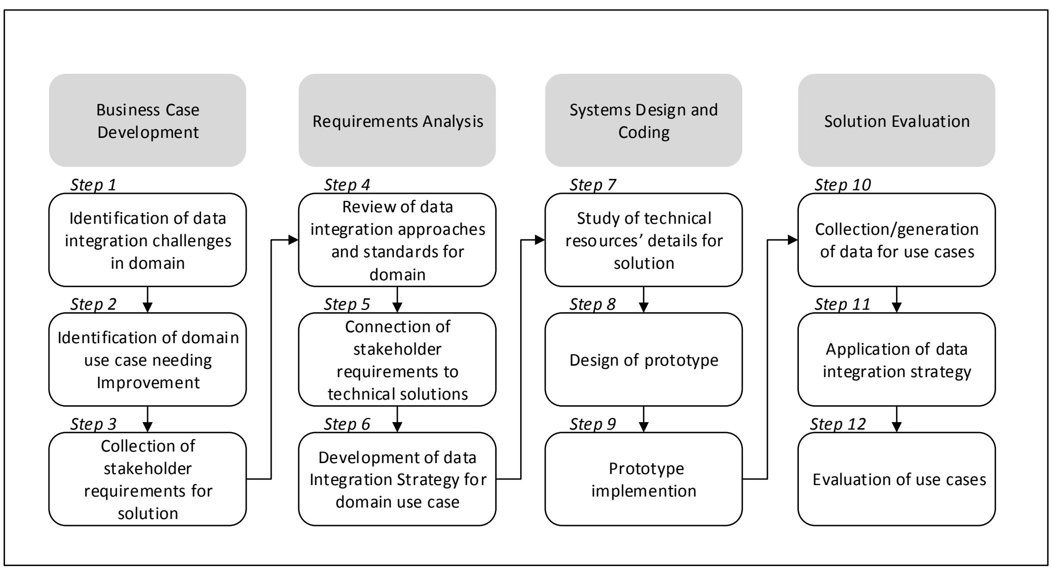


Figure 4.1: Process model for research

4.3 Research Techniques

This research combines quantitative and qualitative research design techniques in order to balance out the limitations of individual methods. Figure 4.2 below highlights the various techniques used for the different process steps and they are described as follows:

Techniques \ Process Steps	1	2	3	4	5	6	7	8	9	10	11	12
Literature Review	☑	☑	☑	☑	☐	☐	☑	☐	☐	☑	☐	☑
Functional Decomposition	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Decision Modelling	☐	☐	☐	☐	☑	☑	☐	☐	☐	☐	☐	☐
Process Modelling	☑	☐	☐	☐	☐	☐	☐	☑	☐	☐	☑	☐
Interviews	☑	☑	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐
Prioritization	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Use Cases and Scenarios	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☑	☐
Ontological Modelling	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☑	☐
Prototyping	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☑	☐
Acceptance and Evaluation Criteria	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑

Figure 4.2: Techniques applied across this research

a) Literature Review: this was first conducted to identify the data integration challenges in the domain (Chapter 2). A second literature search was used to identify domain use cases needing improvement (Chapter 3). Thereafter, another review was conducted to identify applicable research techniques for the study (Chapter 4). In addition, literature search was carried out on data integration techniques to understand the available resources in developing a solution for this research's data integration challenge (Chapter 2). Furthermore, literature search helped to understand the technical resources to be applied in the proposed data integration solution (Chapter 5). Lastly, literature review was carried out to give context to the synthetic data used in one of the solutions that was evaluated (Section 6.2.1).

b) Functional Decomposition: is used to break down complex processes, organisational unit, product scope, or other elements into more feasible components (IIBA, 2015). This enables traceability from high-level concepts to low-level concepts. The generation of simpler constituents allows for a thorough analysis. The appropriate level of functional decomposition depends on where, why, and when to stop decomposing the subject for a given set of analysis objectives. Functional decomposition assumes that parent components are completely described by their sub-components and that each sub-components can only have one parent component in a functional hierarchy (IIBA, 2015). The following are some of the common uses of functional decomposition:

i. For project Work Breakdown Structure (WBS) which decomposes endeavours into phases, milestones, work activities, tasks, work items, and deliverables
ii. For solution detailing to enable understanding of the solution components for reverse engineering purposes
iii. For process detailing to ensure clarity in implementation, modification, optimisation, measurement, and estimation
iv. For decision support, by detailing the underlying models and linked dependencies to ensure the best outcomes.

Functional decomposition results can be expressed as a combination of plain textual descriptions, hierarchical lists, descriptions using special formal notations (e.g. Mathematical formulas and programming languages), and visual diagrams. A wide variety of diagramming techniques are used for functional decomposition. These include:

- work breakdown diagram
- tree diagram
- nested diagrams
- use case diagram
- flow diagrams
- state transition diagram
- cause-effect diagrams
- decision trees
- mind maps
- component diagram
- decision model

Figure 4.3 shows an example of how a function has been decomposed into sub-components using a work breakdown diagram. In this research, functional decomposition is used for decomposing the decommissioning work process to understand the activities involved.

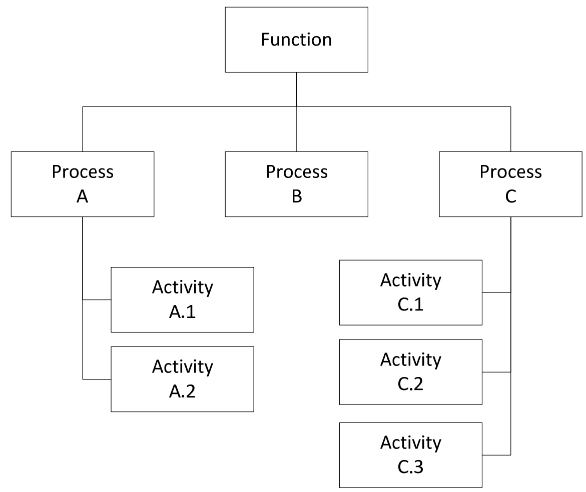


Figure 4.3: Illustration of functional decomposition

c) Decision Modelling: this combines data and domain knowledge to make specific decisions (Pollack, 1963). It can be used for both simple and complex decisions. Decision modelling for simple decisions combines a set of rules with data elements to output a decision outcome. On the other hand, decision modelling for complex decisions break down decisions into their individual components so that component parts can be separately described before being combined to make an overall decision. There are multiple approaches to decision modelling, but the two most common techniques are decision trees (IIBA, 2015) and decision tables (Linetzki, 2012). Decision trees combine multiple simple decisions in a network to handle a complex decision. On the other hand, decision modelling using decision tables enable multiple perspectives and rules to be combined in a single table in making basic decisions.

This research makes use of decision tables in making decisions about researched data. A decision table has columns with rules and rows containing values to be evaluated against the rules. This can be interchanged such that rows represent the rules and the columns contain the values. When the conditions of a given set of rules evaluate to true for a set of input data, the outcome or action specified for that rule is selected. Table 4.2 demonstrates a simple example of how decision tables work. The table columns are subjected to rules that seek pumps carrying liquids, having minimum efficiency of 5 and not less than 6 years of age. The column to the far right is the outcome column and indicates the decision. The rows in the table contain specific instances of the corresponding column class. And the data elements being considered are located at the extreme left column of the decision table. When the conditions of the specified rules for the considered data element (pumps) are true, then the specified decision outcome will be positive. In this case pumps C346 and H342 qualify as having efficiency of at least 5 and not less than 6 years of age.

Table 4.2: Sample decision table

	Decision rules			
	≥ 6	= L	≥ 5	
Pump ID	Age (Yrs)	Fluid	Efficiency (hp)	Eligibility
C346	6	L	6	Eligible
C301	7	G	7	Ineligible
B222	8	L	4	Ineligible
H342	9	L	8	Eligible
B242	5	L	5	Ineligible

Decision: Pumps carrying liquid, having minimum efficiency of 5 and not less than 6 years of age

d) Process Modelling: is the analytical representation of an enterprise or system by using standardized graphical model. The graphical model is referred to as a process model (Repa, 2011). Among the various uses of process models are the following:

- definition and documentation of research methods
- documentation of information flow within an enterprise or system
- investigation of the challenges within an enterprise or system and understanding how to improve them
- coordination of elicitation processes with relevant stakeholders
- visual presentation of the future state of a process
- description of existing work process of an enterprise or system
- description of a solution work plan

Process models come in a variety of notations (Aguilar-Savén, 2004; IIBA, 2015). However, they typically contain some or all of the following key elements:

i. Activity: a discrete step or piece of work that is part of a process. This may be further decomposed into subprocesses that have their own process elements. The participant related to a process step may also be linked to it.
ii. Event: a zero-time occurrence that starts, delays, or ends a task within a process or the process itself.
iii. Directional Flow: a path that shows the logical flow of the process steps. This is typically in the direction that text reads.
iv. Decision Point: a point in the process where the workflow divides into two or more paths that are parallels or mutually exclusive alternatives.
v. Link: a connection between process elements.
vi. Role: a person or group with function involved in the process.

Other elements that may be included are data or material inputs and outputs, and call-out descriptions that supplement the graphical representation. Figure 4.4 shows a sample process model where an event triggers Task A in Company Y which produces an output that is combined with the output from Task B by Company Z. The product of that merging undergoes Task C by Company Y and a decision is made based on business rules whether to output the final deliverable or repeat Task C.

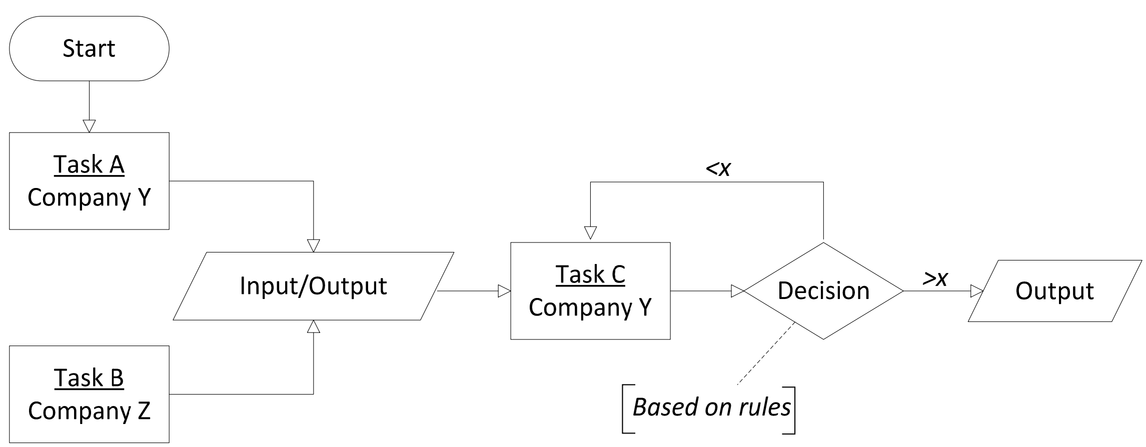


Figure 4.4: Sample process model

Given that this research is *IS* focused Unified Modelling Language™ (UML) activity diagrams and Data Flow Diagrams (DFDs) are used. UML is a notation specified by the Object Management Group for describing software application structure, behaviour, and architecture (IIBA, 2015). DFDs are used for mapping the data motion in a process and to capture related constructs that have bearing on data latency and accessibility (Li and Chen, 2009). DFDs show the source of data, the data processing activities and the storage plan. They illustrate the transformation of data between externals (entities) and processes - the output from one is the input to another. DFDs have multiple layers of abstraction. At the highest level is the context diagram (PMI, 2013) used in this research. This captures a whole system but with a limited level of detail. An advantage of context diagram is the ease with which it can be understood by non-experts. The key elements of a context diagram include:

i. *Externals:* These refer to persons, organisations or systems that produce or store data. They are considered as outside of the system in focus. Externals are represented with rectangles and have data flowing into or out of them, or both.
ii. *Data Store*: This is a repository where data can be stored and read from. The data store is represented with dual parallel lines or an open-ended rectangle. Also, it must have, at minimum, a data flow into it.
iii. *Process*: This helps to transform input data into an output. The process is represented with a circle or rectangle with rounded corners. Each process has, at minimum, a flow into and out of it.
iv. *Data Flow*: This represents the movement of data between externals, data stores and processes. The data flow is represented with a line carrying an arrow in the direction of the flow.

A sample schema of DFD is as shown in Figure 4.5. DFDs are useful for defining the scope of a system. This will help to identify the boundaries of the system and discover duplicate data elements. Also, the important connections within the system can be established in the process.

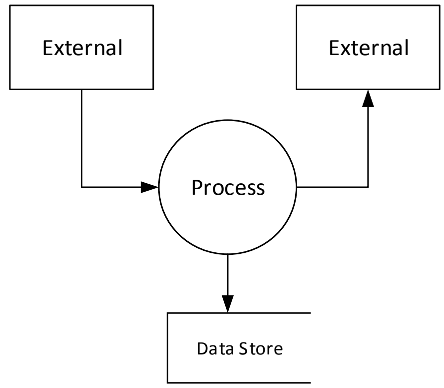


Figure 4.5: Sample Data Flow Diagram

e) Interviews: Unstructured interviews were conducted at multiple workshops on topics related to this research. Unstructured interview is described in Jamshed (2014) as a controlled conversation skewed towards the interests of the interviewer. Sork (1997) describes workshop as a forum for interacting with groups of stakeholders to gain more information about them or activities linked to them. Given that sessions during workshops enables the eliciting of information from participants in informal ways, several industry workshops were attended to engage key stakeholders on their understandings of the current state of decommissioning and what the future needs are. Details on workshops attended are provided in Appendix 2.

f) Prioritization: Prioritization ranks the requirements of a process or a given set of alternatives in order to establish their relative importance within a change context (Vestola, 2010). Prioritizing a requirement or an item improves or reduces its relative importance. This may influence the sequence of its implementation. Inter-dependencies between requirements could also be a basis for prioritization. Prioritization is beneficial because it helps to optimise processes or choose the best option among alternatives. The basis on which prioritization is done is usually agreed by relevant stakeholders considering factors including the following:

   i. *Benefit*: refers to the value that may be added as a result of implementing a requirement in the context of a goal. This may be in form of specific capabilities, strategic goal or quality benchmark.

ii. *Penalty*: refers to the outcome that may result for not implementing a requirement. For example, a regulatory requirement will take precedence over other competing issues as sanctions may be imposed for non-compliance.
iii. *Cost*: refers to the resources needed to achieve a requirement. A need may be downplayed because of astronomical cost. Cost is often coupled with benefit for cost-benefit analysis.
iv. *Risk*: refers to the measure of uncertainty associated with successfully meeting a requirement. The level of risk acceptance will depend on the risk tolerance of the stakeholders. In a situation where a solution is fraught with unknowns, a proof of concept may be commissioned first.
v. *Dependencies*: refers to the relationship between requirements. Some requirements can only be met in a logical sequence. As a result, the ones that need to take place first are prioritized.
vi. *Time Sensitivity*: refers to the time constraint on some requirements. An example is the time-to-market scenario where value outcomes for a product will be exponential if it can be delivered ahead of market competition.
vii. *Stability*: refers the inconclusive nature of some requirements. This may be due to disagreement among stakeholders or the need to further investigate such needs. This consideration results in a lowered priority to avoid rework and waste.

Prioritization can be done by grouping, ranking, budgeting or negotiation of requirements. The negotiation approach played out in this research as it simply involved observing relevant stakeholders at multiple workshops reach consensus on the priority of subjects related to this study.

g) Use Cases and Scenarios: Use case development is a very important step in the design, implementation, analysis and testing process of new systems. They are used to model the desired behaviour of solutions by showing how user interactions with them achieve specific goals. The result of these interactions helps define the expected outcomes of the solutions. A use case can have multiple scenarios describing alternative ways an actor (person or external system) can interact with a solution in order to achieve a goal. Scenarios are written as a series of steps performed by actors or by the solution in order to achieve a goal. They detail different paths that can be followed by defining primary and alternative flows. The primary flow represents the most direct way to accomplish the goal of the use case while special circumstances and *exceptions that result in a failure* are documented in alternative or exception flows (Bittner, 2002).

Use cases are written from the point of view of the actor and avoid describing the internal workings of the solution. As a result, they are easy to comprehend by users. Because use cases capture what a system is supposed to do, they help systems developers conceive designs that align with the users' point of view. Also, because a use case details the sequence of actions a system needs to perform to yield results, they are useful as user documentation when the system becomes operational. This research developed solution for a prioritised use case and presented multiple scenarios through which the goal of potential users can be met. Table 4.3 below shows the essential contents of a use case template and their descriptions. The use case templates developed in this research are provided in Appendix 3.

Table 4.3: Essential contents of a use case template (Coleman, 1998)

Contents	Description
Use Case Identifier	Details on name, reference number, revision date, status, author etc.
Description	Goal to be achieved by use case and sources for requirement
Actors	List of primary and secondary actors involved in use case
Assumptions	Conditions that must be true for use case to terminate successfully
Steps	Interactions between actors and system important for achieving goal
Variations (optional)	Variations in the steps of a use case
Non-Functional (optional)	List of non-functional requirements that the usecase must meet
Issues	List of issues that are yet to be resolved

h) Ontological Modelling: Ontological modelling helps to define a consistent vocabulary for a domain (Spyns, Meersman and Jarrar, 2002). The model generated provides a consistent approach to analysing data and its relationships. Ontological models can be represented in a number of formal languages (Doan, Halevy and Ives, 2012b). However, what is important is that the data elements, their attributes, significant relationships, and constraints are captured in the ontology generated. Figure 4.6 shows a simple example of ontological modelling. In the example, object and datatype properties that represent a modeller's conceptualization of a domain are used to link the constituent data elements. For example, data entities 'Pipe' and 'Class4' are linked by object property 'hasInspectionClass'. This means that the pipe has a Class4 inspection class.

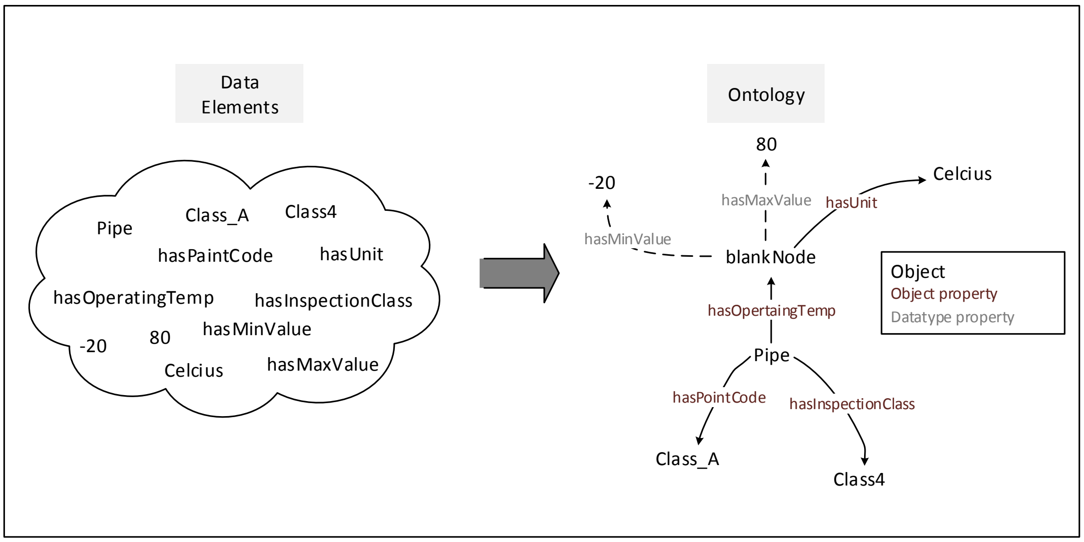


Figure 4.6: Ontological modelling

i) Prototyping: is an established method for product design (IIBA, 2015). It works by providing an early model of the final result, known as a prototype. This is used to elicit and validate stakeholders' needs through an iterative development and testing process (PMI, 2013). Prototypes model the future state options and are used for discovering potential value of solutions. The prototyping process helps to identify missing or improperly specified requirements and assumptions that are not well founded by simulating the performance of a solution at the early stages of design. Prototypes can be non-working models (e.g. scale models in architecture), working representations (e.g. software products), or digital depictions (e.g. simulation with 3D models in engineering). This research is focused on technology application for a process in the process industry. It seeks to prove the value of a solution concept by simulating the relevant business processes with technology tools and comparing it to the existing practices. To develop the prototype for this research, a software development methodology is essential in managing the process.

Rapid Application Development (RAD) method (Despa, 2014) is used for this research. It is a minimalist and agile method that emphasises using reusable components and adopting a more flexible planning approach. The RAD method gives much faster and higher quality results when compared to the traditional software development techniques such as the waterfall model. The waterfall model is rigid and less interactive because progress is in phases (design, planning, development etc.) and unidirectional (downward). On the other hand, RAD imposes less emphasis on planning activities and focuses more on development. In RAD, development cycles are time boxed and multiple cycles can be run at the same time. Feedback is given after each module development is completed. The feedback validates the solution or requests further modification. Thereafter the completed modules are combined. RAD is suitable for all types of project - large and small. Figure 4.7 describes the RAD process and the following discusses its life cycle stages relevant to his research:

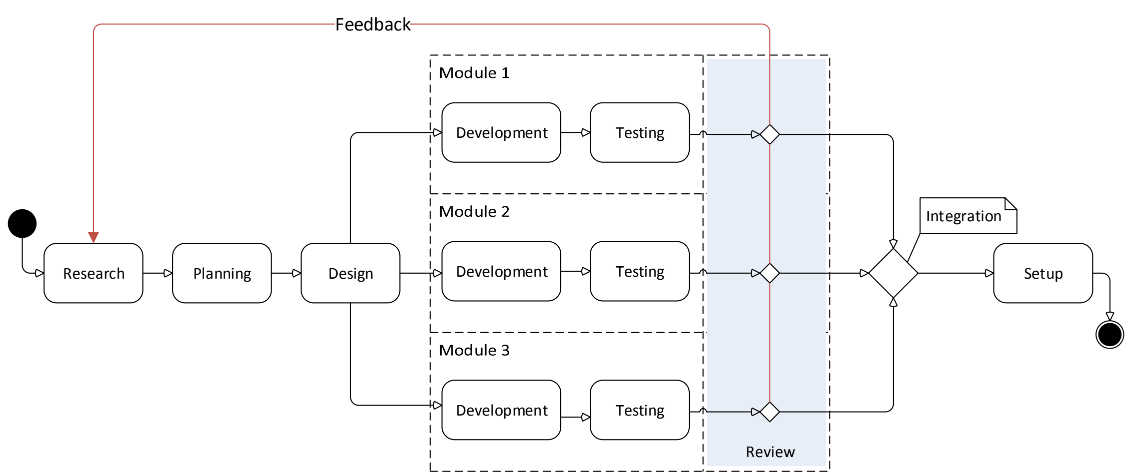


Figure 4.7: Rapid Application Development (RAD) process

i. *Research* is the stage where the project stakeholders investigate what the appropriate solution for a problem should be. The solution requirements are formulated based on the project goal and evaluated to ensure their fitness for purpose. Also, the appropriate technology resources (frameworks, Application Programme Interface [API], libraries, versioning tools and hosting infrastructure) required to build the software solution are investigated.

ii. *Planning* is the stage where the software development work is broken into smaller tasks that can be easily managed by human resources. For each task, the required quality and duration are associated. The tasks are also sequenced in a logical order and appropriate technologies matched to them. Lastly, a baseline is firmed up for monitoring the progress of the project.

iii. *Design* is the stage where the overall architecture of the software application to be developed is defined. The hardware and systems requirements are also specified. These specifications are used as input for implementing the software solution.

iv. *Development* is the stage where the computer codes are generated, and the software solution is built. Here, the development and testing environments for the codes are set up. Also, it is during this stage that debugging of the computer codes starts.

v. *Testing* is the stage where errors related to computer programming and design are identified and recommended for fixing. Programming errors may make the application crash or behave in a way that is in variance to the designed architecture. They also can result in security or usability issues. Design errors on the other hand are variations between the specified requirements for the solution and the solution produced.

vi. *Integration* is the stage that combines the various developed modules into one solution. It takes place after the validation of the test results of the completed modules, and its function is to ensure that the developed modules work together as a whole.

vii. *Setup* is the final stage relevant to prototyping in this research. It entails configuring the required hardware and software resources. Software setup in this case includes copying the source code, configuring the database, installing third party applications and configuring APIs. Once these processes are completed, the solution is passed through another round of testing. Thereafter content is added to the application and desired evaluation is carried out.

j) Acceptance and Evaluation Criteria: acceptance criteria describe the minimum requirements that must be fulfilled for a solution to be accepted by stakeholders. They are generally expressed as pass or fail when verified against specified benchmarks and are typically used when only one possible solution is being considered. Evaluation criteria on the other hand refer to parameters used to assess a set of requirements in order to choose between alternative solutions. Each evaluation criterion has a scale and its assessment measurements allow for the ranking of solutions based on their value to stakeholders. Expert judgment is applied when there is a need to evaluate a criterion that cannot be measured directly (IIBA, 2015).

Acceptance and evaluation criteria are measures of value attributes that allow for objective and consistent assessment of solutions. By implication, value attributes are the characteristics of a solution that inform stakeholders of its worth. They represent the qualities that should be present or absent in a solution. Examples include:

i. ability to generate information in a specified format,
ii. ability to perform or support specific functions,
iii. extensibility and scalability of a solution,
iv. precision of results, and
v. reliability against power disruption.

Table 4.4 shows the characteristics of the two classes of value attributes. Validating or measuring value attributes may sometimes involve designing tools and instructions for performing the assessment, as well as for processing and storing the results. In order to assess a solution against acceptance or evaluation criteria, the solution needs to be constructed in a measurable format. When evaluating multiple solutions, those with better performance and lower costs are ranked higher. This research will develop a solution prototype and test it against acceptance criteria that are based on the elicited requirements. Using the concept of the prototype, solutions for two use cases will be developed and evaluated by tests. In one set of tests, custom instructions will be used for evaluation. In the second, precision and recall evaluation method (Fawcett, 2006) will be used.

Table 4.4: Characteristics of value attributes

Value Attribute	Solution	Assessment	Validation
Acceptance Criteria	Single	Benchmarking	Pass/Fail
Evaluation Criteria	Multiple	Perfomance Measurement	Rank

Precision and recall are based on the confusion matrix. The confusion matrix is a table used for describing the performance of a classifier (person or system) on a set of test data with known values (Davis and Goadrich, 2006). Figure 4.8(A) describes the data set in a binary classification case as in this research. The data set has P positive values and N negative values. A prediction of this composition by a classifier will result in correctly identified values that sum up to TP (true positives) and wrongly identified values that sum up to FP (false positives). Similarly, the unpredicted positive values will sum up to FN (false negatives) and unpredicted negative values TN (true negatives). This distribution is summarised in the confusion matrix in Figure 4.8(B). To compute the precision of the classifier, the number of true positives is divided by the sum of true positives and false positives as in Equation 4.1. To compute recall the number of true positives is divided by the sum of true positives and false negatives as in Equation 4.2.

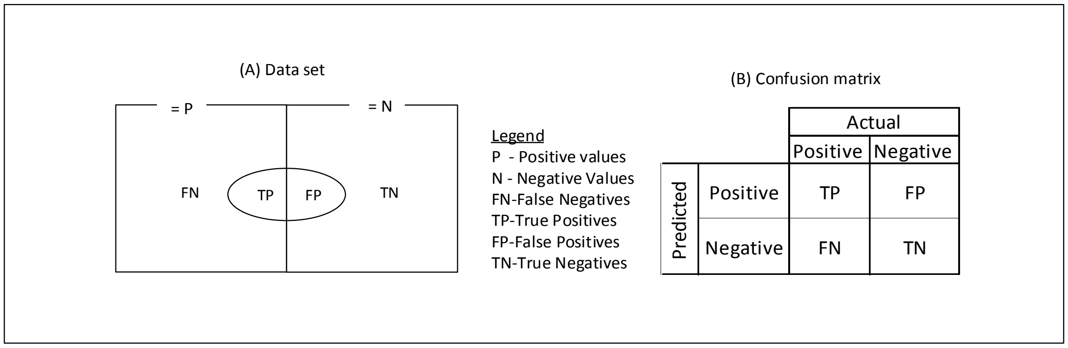


Figure 4.8: Confusion matrix for a binary classification case

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{4.1}$$

$$Recall \ = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{4.2}$$

Precision and recall are useful for measuring the success of prediction when the classes in a data set are very imbalanced – not evenly distributed. These metrics are often applied in evaluating query answering systems (Soibelman *et al.*, 2008; Tserng *et al.*, 2009). Consequently, appropriate for this research. Precision measures the magnitude of useful results of a query output, while recall measures the completeness of the useful output relative to the source. In essence, precision measures quality while recall measures quantity. In the context of information retrieval good precision returns substantially more relevant results than irrelevant ones; and good recall returns most of the relevant results in the data set (Davis and Goadrich, 2006). Perfect precision (score of 1.0) means that the complete output of a search is relevant but says nothing about whether all relevant data points were retrieved. On the other hand, perfect recall (score of 1.0) means that all relevant data points are returned by a query but says nothing about the irrelevant ones also retrieved.

In order to evaluate the effectiveness of a classifier, its precision and recall are considered. Figure 4.9 shows precision-recall curves for representing the trade-off between precision and recall at different thresholds of two classifiers. The area under a curve is a single performance measure, for precision and recall, of a classifier. By inspection, the area below the random classifier in Figure 4.9 is greater than that of classifier 1 in the same graph. This implies that the random classifier outperforms classifier 1 in general. However, given that precision and recall are compared for a fixed level of each other, (e.g. precision at a recall level of 0 is 1), classifier 1 can be said to outperform the random classifier at low levels of recall. This is because a classifier with high precision, but low recall, returns mostly correct results. Conversely, a classifier with high recall, but low precision, returns mostly incorrect results. An ideal classifier will aim for both high precision and high recall.

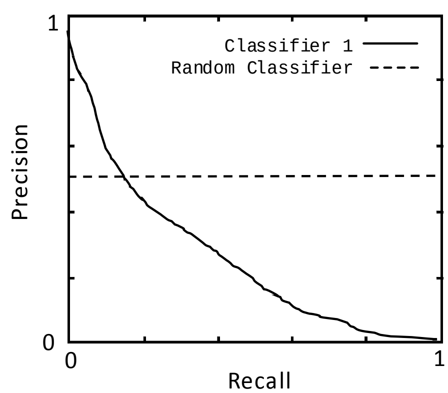


Figure 4.9: Precision-recall trade-off

Other ways to combine precision and recall is to compute their weighted harmonic mean, the F-measure, or the geometric mean, the Matthews Correlation Coefficient (Sing *et al.*, 2005). The F-measure is the more commonly used of the two and is adopted for this research. The general formula for the F-measure is given in Equation 4.3. The value of β in the Equation is the ratio of recall to precision and measures the importance a user attaches to the two variables. The F_1 score is a specific example of the general F-measure ($F_{\beta)}$ and is a scenario where equal weight is applied to both precision and recall in achieving an optimal blend. The formula for F_1 is given in Equation 4.4. The higher the F_1 score the better the classifier performance.

$$F_\beta = (1 + \beta^2) \, \frac{Precision \, . \, Recall}{\beta^2 \, . \, Precision \, + \, Recall} \tag{4.3}$$

$$F_1 = 2 \, \frac{Precision \, . \, Recall}{Precision \, + \, Recall} \tag{4.4}$$

Lastly, the main visualisation technique for showing the performance of a classifier is the Receiver Operating Characteristic (ROC) curve. The ROC curve summarizes the performance of a classifier, in terms of its precision-recall trade-offs, over all possible thresholds. It is generated by plotting the True Positive Rate against the False Positive Rate as the threshold for assigning observations to a given class is varied. True Positive Rate of a classifier (also called sensitivity) is equal to its recall value as in Equation 4.2. To compute the False Positive Rate (also called fall-out), the number of false positives is divided by the sum of false positives and true negatives as in Equation 4.5. Figure 4.10 shows an example of a graph with multiple ROC curves. The area under a curve is a single overall performance measure, for precision and recall, of a classifier. By inspection, the area below classifier 1 in Figure 4.10 is greater than that of the random classifier in the same graph. This implies that classifier 1 outperforms the random classifier in general.

$$False\ Positive\ Rate = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{4.5}$$

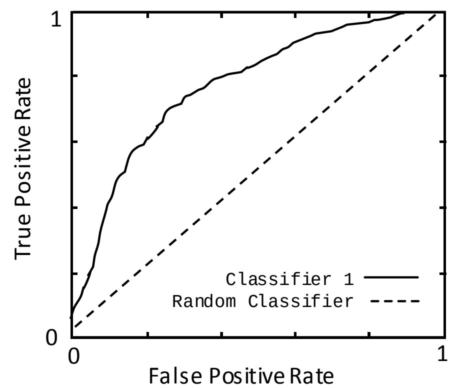


Figure 4.10: Receiver Operating Characteristic (ROC) curve

4.4 Summary

This chapter described the research design in detail and explained all the techniques applied throughout the research. The next chapter describes the design and testing of an ontology-based data integration solution for process plant data.

Chapter 5

System Design and Coding

5.1 Requirements Analysis

Several studies (Wiesner et al. (2011); Mutis and Issa (2012); and Bilal, Oyedele, Qadir, et al. (2016)) show evidence of the need for data integration in built environment projects and justify the variety of solutions currently available. Using this knowledge and engaging process industry stakeholders (Appendix 2), the following requirements are underscored as being essential for any data integration solution.

a) Data migration**:** effort and time required to extract data from different sources are factors in the choice of tools or technology framework for data integration (Rahm and Do, 2000). A recent survey of data scientists indicated that they spend 60% of their time cleaning and organising data and nearly 60% of them consider this exercise the least enjoyable of their tasks (CrowdFlower, 2016). Consequently, a straightforward process that allows data in the sources to be transformed in a reasonable amount of time is beneficial.

b) Data quality checks: being able to check for inconsistencies of the integrated data is very important to the technology users because they do not want to have spent a lot of effort integrating data to then find out in the end that there are errors. As a result, the introduction of quality check measures in the data processing cycle is considered valuable in any tool or technology framework that enables data integration.

c) Reuse and Extendibility: serious care is required for every data integration task to avoid errors as it takes time to complete. Consequently, any solution that allows reuse of data mapping and query files multiple times is very desirable. Also, the possibility of extending the integrated data without errors is considered as valuable.

5.2 System Design

Dealing with decommissioning waste requires an ability to integrate life cycle information for aging assets (DNS, ZWS and ABB, 2015). This type of data integration effort requires ingestion of data in a variety of structures, resolution of overlapping terminologies and the ability to query data across sources. These are key functionalities Curé and Blin (2015) highlighted about Semantic Web technologies, and specified in the Part 8 of ISO 15926. The Semantic Web uses Resource Description Framework (RDF) (Cyganiak, Wood and Lanthaler, 2014) as a standard graph-based data model with globally defined identifiers and terminologies. This approach supports bringing together disparate data sources and quickly highlights inconsistent terms. To map data (i.e. create an ontology), Doan et al. (2012c) describe that the contents of the sources are associated with semantic markups to form identifiers which facilitate easier integration and more accurate search results.

The proposed framework is based on the consideration that engineers are not computer scientists but are capable of understanding and making use of this technical knowledge. While engineers may not be able to build comprehensive or perfect ontologies, they can create functional ontologies that will serve their information management purposes. Figure 5.1 shows the conceptual framework for the ontology-based work process for data integration. The steps in the process are discussed below:

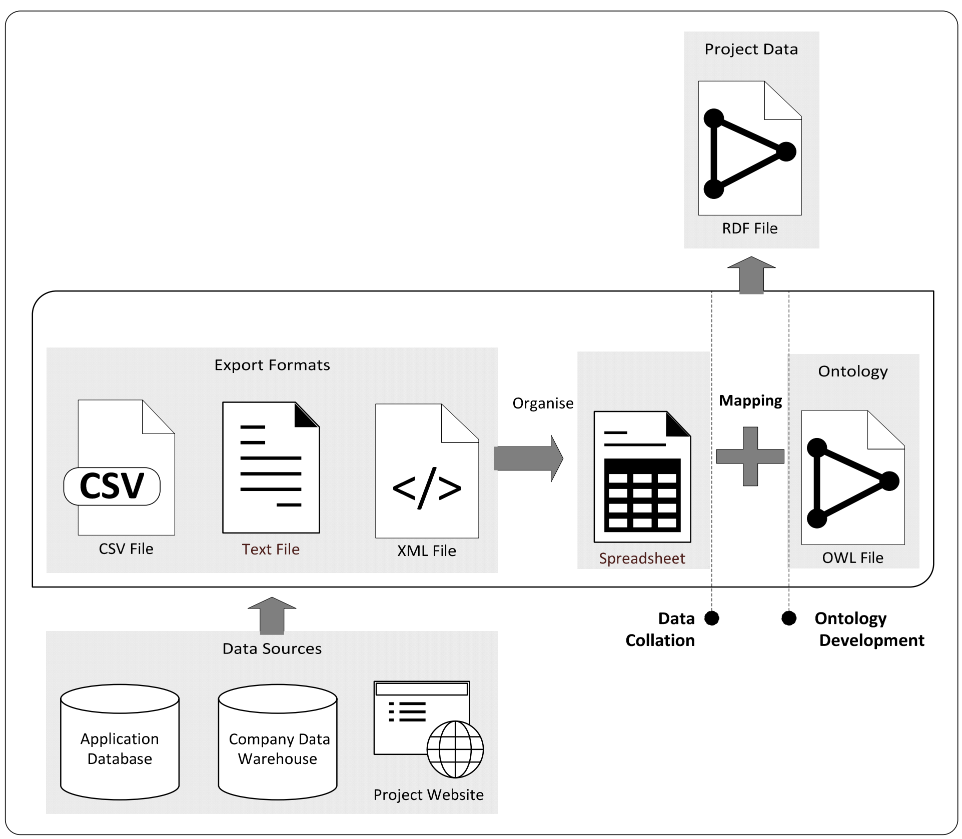


Figure 5.1: Integration framework for ad hoc data projects

a) Data Collation: Soibelman *et al.* (2008) list some of the typical data sources in built environment projects. These sources include structured data (e.g. relational databases and spreadsheets); semi-structured documents (e.g. JSON and XML); unstructured text documents (e.g. contracts, specifications, change orders, requests for information, meeting minutes, e-mail messages, webpages); and other unstructured multimedia data (e.g. 2D/3D drawings, pictures, audios, videos). With data extraction techniques (Rusu *et al.*, 2013), these sources can be transformed into preferred structured formats (Groves, 2016; Roman *et al.*, 2017). However, the data cleaning effort required depends on the status of the source, which is highly variable. Given the variability of the sources, this research is not focused on the data cleaning process. It is premised on the notion that most of the sources can be exported or organised into a spreadsheet document. Spreadsheets are favoured for this data integration task because of their widespread use in engineering (Lee *et al.*, 2016); and multiple tools, e.g., MappingMaster (O'Connor et al., 2010) and dot15926 (TechInvestLab.ru, 2013), support mappings between them and ontologies. This step is very important because the spreadsheet templates generated will be repeatedly used for updates on a project. In addition, they can be used for similar projects.

b) Ontology Development: Individuals developing ontologies often have different conceptualisations of a given domain, and all may be right, as there is no single right way of modelling a domain. Consequently, it makes sense that for a small data project, an ontology based on the understanding of the data sources, by the users, is what is developed. Also, it is easier to interpret, accommodate and model what is rather than trying to fit reality to a standard ontology (Rezgui *et al.*, 2011). To develop the required ontology, a list of entities in the sources is first generated. Thereafter, classes are identified and sorted into taxonomy of superclasses and subclasses where applicable. Attributes and relations are also identified. An ontology of the entities in the data sources, as documented in spreadsheets, is then created, and axioms defined. Where there is a standard ontology for the domain, its standard entities should be abstracted for the required purpose. Also exiting templates from a standard ontology that meet the requirement of a task should be taken advantage of.

c) Data Mapping: This is carried out to resolve the terminology heterogeneity in the data sources. Once the data sources are organised into spreadsheets and the project ontology developed, mappings are created to migrate the data in the spreadsheets into the ontology. The output data is the integrated project data. Mappings used can be generated using the RDF Mapping Language (RML) (Dimou *et al.*, 2014) or any open source library that supports mapping from spreadsheets into OWL. An important advantage of this step is that it lends itself to self-checks. Any errors in the integrated data can be easily corrected by revising the mapping rules and re-importing the source data into the base ontology.

5.3 Software Study

Considering that the Rapid Application Development (RAD) method is the prototyping method adopted for this research, the pool of open source software provides ample resources for solution development. Open source software is developed in a public collaborative manner and shared with a license that allows the users to reuse, modify and redistribute them without limitations (Marsan, Paré and Beaudry, 2012). Open source software is also free and can be easily downloaded from the Internet (Dabbish *et al.*, 2012). According to Hauge et al. (2010), adoption of these products can be either by directly applying them for tasks or by using the products in software development. They identified software development activities to include: making use of open source software development platforms, extending a software application, integrating software applications, adopting open source software development practices, and participating in open source software development communities. The two scenarios that are applicable in this research include using open source software as-is and integrating open source software applications into a system. Challenges associated with open source software relate to the cost of learning and understanding the components, and the time it takes to integrate them (Chen *et al.*, 2008). However, these issues are mitigated by adopting open source software with active community support and useful documentation (Sarrab and Rehman, 2014). Table 5.1 describes the open source software resources applied in this research.

Table 5.1: Research software resources

Software	Function (*Plugin*)	Description	References
Protégé	Ontology development SPARQL Querying Reasoning Data mapping (*Cellfie*) Graphical visualisation(*VOWL*) Ontology-Based Data Access (*Ontop*)	Protégé is an open source tool used for creating OWL ontologies. Protégé has multiple plugins that allow it to carry out functions like SPARQL querying, reasoning and data mapping. In this research, Cellfie, a MappingMaster plugin to Protégé is used for data mapping, and VOWL used for depicting elements of the ontology. Also, Ontop, another Protégé plugin is used for querying relational data sources provided in terms of the ontology to which they are mapped.	Musen (2015), Horridge et al. (2011), Protégé Project (2017), O'Connor et al. (2010), and Calvanese et al. (2017)
Twinkle	SPARQL Querying GUI	Twinkle is an open source graphical user interface (GUI) for querying RDF data. It can be used for querying local and remote RDF data sources.	Dodds (2007)
dot15926 Editor	Ontology development Data mapping ISO 15925 Referencing	dot15926 Editor is an open source architecture and set of specific libraries that enable working with ISO 15926. The Editor is the equivalent of Protégé, except that it is targeted for handling ISO 15926 data. dot15926 Editor enables exploration of existing reference data sources; helps to validate reference data; can be used for developing new reference data; and enables mapping of data from sources.	TechInvestLab.ru (2013)
RDF4J	Triplestore SPARQL endpoint Reasoner	RDF4J is an open source Java framework for storing, inferencing and querying RDF data. RDF4J can be connected to local and remote triplestores using its easy to use API. The RDF4J framework supports SPARQL querying and multiple RDF serializations including Turtle, RDF/XML, N-Quads, JSON-LD, N-Triples, TriG and TriX. The framework also allows the representation of business rules using SPARQL Inferencing Notation (SPIN).	Eclipse RDF4J (2017) and Knublauch (2013)

5.4 System Implementation

The test use case (use case 1) is about pipe spools, which are common components in large oil and gas construction projects. Activities involved in pipe spool fabrication include cutting, fitting, welding, quality inspection, non-destructive testing, post heat treatment and painting (Standards Norway, 1996). In the fabrication of pipe spools, raw pipes and pipe fittings (e.g. elbows, flanges, reducers) are welded according to a design. To achieve this, piping isometrics (iso) from the design are divided into spools that can be fabricated. These sub-assemblies are tacked, welded and checked for quality of work. Additionally, x-ray tests are carried out to certify the welds, and leak tests completed to check for fluid leakage. This demonstration case is adopted from a pipe spool fabrication project, conducted by Ariosh (2018) - an engineering service contractor, having challenges with work progress information.

Information on the progress of work exists in multiple records and is making the determination of the overall completion status of each spool piece inconclusive. Information sources, as shown in Figure 5.2, include a line list export from a 3D design application (iso number, service, line size); spool status report from the project engineer (iso number, Spool number, completion status); leak test report from a service contractor (Type, spool number, completion status); and x-ray test report from another service contractor (spool number, completion status). The x-ray of welded joints precedes pressure tests in the work process, and reports from the test activities are used to update the project engineer's overall project status.

A problem with the data is that there are gaps in two of the reports – overall and x-ray. The affected reports do not represent the most up-to-date situation at the fabrication yard because some of the spools have had to be retested. Another problem is that stakeholders have represented iso numbers differently e.g. '18"-P-A12-002 001-B' in the pressure test report is represented as '18"-P-A12-002 001B' in the x-ray report and split into '18"-P-A12-002 001' and 'B' in the overall project report. In carrying out a gap analysis of the spools, it is important to use all the reports to resolve the differences, a process that needs to be repeated throughout the duration of the project. On a typical fabrication project, there could be thousands of spools (Soleimanifar, 2016) and it will be very inefficient and error-prone to manually retrieve the correct status of each spool by looking through each report. As a result, for such an ad hoc data project, this ontology-based data integration work process is appropriate.

Line List from 3D Export

Iso number	Service	Line size (")
18"-P-A12-002 0001	Process	18
16"-FL-A12-002 0001	Flare	16
2"-SV-A12-002 0006	Shut Down	2

Overall Project Status

Iso number	Spool number	Completion status
18"-P-A12-002 0001	A	Not Started
	B	Ongoing
	C	Ongoing
16"-FL-A12-002 0001	A	Ongoing
	B	Ongoing
2"-SV-A12-002 0006		Completed

Pressure Test Report

Type	Spool number	Completion status
Hydro	18"-P-A12-002 0001-B	Not Started
Hydro	18"-P-A12-002 0001-C	Ongoing
Hydro	18"-P-A12-002 0001-A	Ongoing
Air	2"-SV-A12-002 0006	Ongoing
Hydro	16"-FL-A12-002 0001-A	Ongoing
Hydro	16"-FL-A12-002 0001-B	Ongoing

X-Ray Report

Spool number	Completion status
18"-P-A12-002 0001B	Not Started
18"-P-A12-002 0001C	Not Started
18"-P-A12-002 0001A	Not Started
2"-SV-A12-002 0006	Not Started
16"-FL-A12-002 0001A	Ongoing
16"-FL-A12-002 0001B	Ongoing

Figure 5.2: Sample data about 6 spools from different construction project sources

To address the above stated problems, the data sources are pre-processed and an ontology of all the available data is created. Thereafter, available data from the sources are mapped to the appropriate entities in the ontology. By doing this, a uniform representation of the data sources is generated, and queries can be used to check for consistency and extract information of interest - including the most up-to-date status of each spool piece. Following the steps of the conceptual framework, the first step is to collate the data. The pre-processed data are as presented in Figure 5.3 and include annotations identifying the source of each data point recorded under the header ‘Issue’. Also, iso numbers in the overall project report are fully completed to ensure each row has a value.

Line List

	A	B	C	D
1	Iso number	Service	Line size	Issue
2	18-P-A12-002 0001	Process	18	Line List
3	16-FL-A12-002 0001	Flare	16	Line List
4	2-SV-A12-002 0006	Shut Down	2	Line List

Poject Report

	A	B	C	D	E
1	Iso number	Spool number	SpoolNumber	Completion status	Issue
2	18-P-A12-002 0001	A	18-P-A12-002 0001A	Not Started	Project Report
3	18-P-A12-002 0001	B	18-P-A12-002 0001B	Ongoing	Project Report
4	18-P-A12-002 0001	C	18-P-A12-002 0001C	Ongoing	Project Report
5	16-FL-A12-002 0001	A	16-FL-A12-002 0001A	Ongoing	Project Report
6	16-FL-A12-002 0001	B	16-FL-A12-002 0001B	Ongoing	Project Report
7	2-SV-A12-002 0006		2-SV-A12-002 0006	Completed	Project Report

X-Ray Report

	A	B	C
1	Spool number	status	Issue
2	18-P-A12-002 0001B	Not Started	X-Ray Report
3	18-P-A12-002 0001C	Not Started	X-Ray Report
4	18-P-A12-002 0001A	Not Started	X-Ray Report
5	2-SV-A12-002 0006	Not Started	X-Ray Report
6	16-FL-A12-002 0001A	Ongoing	X-Ray Report
7	16-FL-A12-002 0001B	Ongoing	X-Ray Report

Pressure Test Report

	A	B	C	D
1	Type	Spool number	Completion status	Issue
2	Hydro Test	18-P-A12-002 0001-B	Not Started	Pressure Test Report
3	Hydro Test	18-P-A12-002 0001-C	Ongoing	Pressure Test Report
4	Hydro test	18-P-A12-002 0001-A	Ongoing	Pressure Test Report
5	Air Leak Test	2-SV-A12-002 0006	Ongoing	Pressure Test Report
6	Hydro Test	16-FL-A12-002 0001-A	Ongoing	Pressure Test Report
7	Hydro Test	16-FL-A12-002 0001-B	Ongoing	Pressure Test Report

Figure 5.3: Processed data in spreadsheets

The second step is to create the ontology for the domain. The generated ontology shown in Figure 5.4 is developed using the ontology design tool, Protégé (Musen, 2015). The development process involves inspecting the data sources and identifying the concepts within them. This reveals that multiple stakeholders are involved in the reporting process. Also, concepts in the domain include spool, service, progress report and completion status. By combining these, it can be deduced that spools in the domain have completion statuses which must be one of 'NotStarted', 'Ongoing' and 'Completed'. The spools are used for some type of services and their completion statuses are reported in progress reports. The progress reports can be grouped into two classes, project and subcontractor reports. The subcontractor report can be classified into pressure test and x-ray reports. The pressure tests carried out are reported in the pressure test report and can be either air leak test or hydro test.

In OWL, classes are assumed to overlap. Consequently, it is important to make classes sharing a superclass, e.g. 'HydroTest' and 'AirLeakTest', disjoint from one another. Applying disjoint classes will ensure that an individual cannot be an instance of more than one of the classes. Also required for the data integration process are the assertions about relationships between concepts. For example, 'Spool is a thing that has completion statuses for pressure test, x-ray test and the overall project as one of 'NotStarted', 'Ongoing' and 'Completed'; and 'Spool has datatype properties, 'hasLineSize' and 'hasLineNumber', with string literals'. Value partition (Horridge *et al.*, 2011), an established ontology design pattern used for restricting the range of possible values to a given list, is applied to 'CompletionStatus' to restrict its values to one of 'NotStarted', 'Ongoing' and 'Completed'. This is enforced by the disjoint between the classes. The covering axiom simply says 'NotStarted', 'Ongoing' and 'Completed' are subclasses of 'Completion Status'; and 'CompletionStatus' is a subclass of the union of 'NotStarted', 'Ongoing' and 'Completed'. These assertions are illustrated in Figure 5.4.

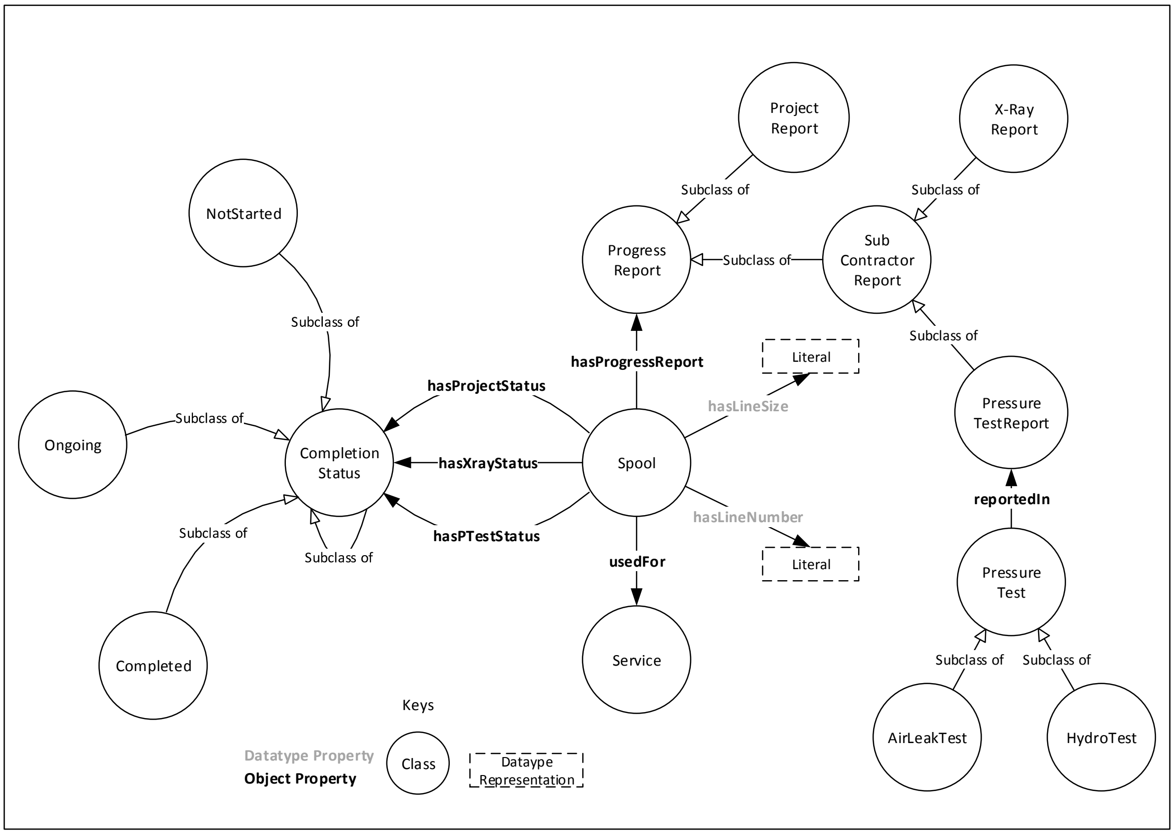


Figure 5.4: Spool ontology

The third and final step is to map the prepared data in the spreadsheets to the created ontology. This is done using the MappingMaster plugin to Protégé called Cellfie (Kaur and Aggarwal, 2017). An important consideration in the mapping process is ensuring the correct OWL constructs are used in mapping the data. For Cellfie, the transformation rules generated are based on OWL Manchester syntax (Horridge and Patel-Schneider, 2012). Figure 5.5 shows the transformation rules for data in the sources. The mapping for the x-ray report, for example, includes axioms describing elements in columns C and B (Figure

5.3) as subclasses of 'CompletionStatus' and 'SubcontractorReport' respectively. In addition, elements in column A (Figure 5.3) are described as OWL individuals having OWL type 'Spool'. These individuals have OWL fact 'hasXrayStatus', a spool property shown in Figure 5.4, with values in column B of the x-ray report (Figure 5.3). The last rule for mapping x-ray report simply adds a comment about the source of entities in the mapping, stating that the latter individuals have a source with value in column C of the x-ray report (Figure 5.3). This described technique is used to map all the spreadsheet sources. In a situation where there is a relation between entities in different sources (spreadsheets), the appropriate reference notation for the open source tool should be explored. An example in this case is the mapping rules for pressure test report, shown in Figure 5.5. The fourth line of the transformation rule describes elements in column B of pressure test report (Figure 5.3) as OWL individuals that are the same with elements in column A of x-ray report (Figure 5.3). Once the transformation rules for the sources are completed, data from the different spreadsheets can be added to the ontology to generate an integrated data set.

X-Ray Report	Class: @C* SubClassOf: SubcontractorReport Class: @B* SubClassOf: CompletionStatus Individual: @A* Types: Spool Facts: hasXrayStatus @B* Annotations: hasSource @C*	Line List	Individual: @B* Types: Service Facts: hasLineNumber @A*, hasLineSize @C* Annotations: hasSource @D*
Pressure Test Report	Individual: @B* Types: Spool Facts: hasPTestStatus @C* SameAs: @'X-Ray Report'!A* Annotations: hasSource @D* Class: @D* SubClassOf: SubcontractorReport Class: @C* SubClassOf: CompletionStatus Class: @A* SubClassOf: PressureTest, usedFor value @B*	Project Report	Class: @D* SubClassOf: CompletionStatus Individual: @C* Types: Spool Facts: hasProjectStatus @D* Facts: hasLineNumber @A* Annotations: hasSource @E* Class: @E* SubClassOf: ProgressReport

* refers to all data in referenced spreadsheet column

Figure 5.5: Data transformation rules for mappings to ontology

To ask questions of the integrated data, SPARQL queries can be parsed via open source graphical user interface (GUI) tools like Twinkle (Almendros-Jiménez, Becerra-Terón and Cuzzocrea, 2017) and YASGUI (Rietveld and Hoekstra, 2014). However, in order to demonstrate the benefit of adding rules to data, the case study integrated data is deposited in an open source triple store called RDF4J (Knap *et al.*, 2018). RDF4J is an open source Java framework for storing, inferencing and querying RDF data. The following query exercises demonstrate information retrieval with and without inferencing.

a) Query without Rules**:** The first two queries in Figure 5.6, Q1 and Q2, extract distinct instances of spools and the progress statuses recorded for each of them in results R1 and R2. Q1 simply says select distinct spools, their x-ray statuses, their pressure test statuses and their statuses in the overall report where RDF triple patterns about them match RDF triple pattern listed in the 'where' clause. In Q2, inspection of the sources shows that a spool is being represented differently across reports. As a result, ?y and ?spool are used to identify the versions. Q2 simply says select distinct spool (?spool version), their x-ray statuses, their pressure test statuses and their statuses in the overall report where RDF triples about them match RDF triples listed in the 'where' clause. These views reveal the inconsistencies present in the current record and creates opportunity for resolving them. Q3 counts the number of spools in the pressure test report to be 6 - shown in result R3.

```
PREFIX spool: <http://www.spools.com/ontologies/spools.owl#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xml: <http://www.w3.org/XML/1998/namespace>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

Q1

```
# Selects spool statuses across reports where equivalence is asserted
SELECT DISTINCT ?spool ?XRay ?PTest ?Project
WHERE{?spool rdf:type [ rdf:type owl:Restriction ; owl:onProperty spool:hasPTestStatus ;owl:someValuesFrom ?PTest].
?y rdf:type [ rdf:type owl:Restriction ; owl:onProperty spool:hasProjectStatus ;owl:someValuesFrom ?Project].
?y rdf:type [ rdf:type owl:Restriction ; owl:onProperty spool:hasXrayStatus ; owl:someValuesFrom ?XRay].
?spool owl:sameAs ?y
}
```

R1

spool	XRay	PTest	Project
18-P-A12-0020001C	Not Started	Ongoing	Ongoing
18-P-A12-0020001B	Not Started	Not Started	Ongoing
18-P-A12-0020001A	Not Started	Ongoing	Not Started
16-FL-A12-0020001B	Ongoing	Ongoing	Ongoing
16-FL-A12-0020001A	Ongoing	Ongoing	Ongoing

Q2

```
# Selects spool statuses where spools have singular ID across reports
SELECT DISTINCT ?spool ?XRay ?PTest ?Project
WHERE{?spool rdf:type [ rdf:type owl:Restriction ; owl:onProperty spool:hasPTestStatus ;owl:someValuesFrom ?PTest].
?y rdf:type [ rdf:type owl:Restriction ; owl:onProperty spool:hasProjectStatus ;owl:someValuesFrom ?Project].
?y rdf:type [ rdf:type owl:Restriction ; owl:onProperty spool:hasXrayStatus ; owl:someValuesFrom ?XRay].
FILTER (?spool=?y)
}
```

R2

spool	XRay	PTest	Project
2-SV-A12-0020006	Not Started	Ongoing	Completed

Q3

```
# Counts the number of spools in the pressure test report
SELECT (count(?spool) AS ?count)
WHERE {?spool rdf:type [ rdf:type owl:Restriction ; owl:onProperty spool:hasPTestStatus ;owl:someValuesFrom ?PTest].}
```

R3

count
6

Figure 5.6: Sample queries on integrated data

b) Query with Rules: An advantage of using an ontology is that the technology has the capability to reason about the data and the rules added to it logically (Horrocks and Patel-Schneider, 2011). Attempts to make use of the latter results would require the user to mentally resolve the logical inconsistency in the output even though queries Q1 and Q2 extracted the semantically reconciled statuses from the reports. To resolve this issue, rules are added to the repository and queries with inference, as shown in Figure 5.7, are run. SPARQL Inferencing Notation (SPIN) (Knublauch, 2013) is used to represent the rules. SPIN is supported by RDFJ4, so the rules are added to the triple store in addition to the project data. Rule 1 simply asks that every spool with pressure test status as ‘Ongoing’ should have x-ray status as 'Completed' and overall project status as 'Ongoing'. Pressure test report is used as the benchmark because it is the only report with up-to-date status. Consequently, when it has a status 'ongoing', x-ray status should be 'Completed', and overall project status should duplicate its status. Rule 2 follows a similar construction, and query Q4 returns the true status of every spool.

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix sp: <http://spinrdf.org/sp#>.
@prefix spin: <http://spinrdf.org/spin#>.
@prefix spool: <http://www.spools.com/ontologies/spools.owl#>.
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

Rule 1
# every spool with pressure test status as 'ongoing' has xray status as 'completed' and project status as'ongoing'.
spool:Spool a owl:Class ;
 spin:rule [ a sp:Construct ;
 sp:text"""PREFIX spool: <http://www.spools.com/ontologies/spools.owl#>
    CONSTRUCT { ?this rdf:type [ rdf:type owl:Restriction ; owl:onProperty spool:trueXrayStatus ; owl:someValuesFrom spool:Completed],
                                          [ rdf:type owl:Restriction ; owl:onProperty spool:truePTestStatus ; owl:someValuesFrom spool:Ongoing] ,
                                          [ rdf:type owl:Restriction ; owl:onProperty spool:trueProjectStatus ; owl:someValuesFrom spool:Ongoing].}
    WHERE {?this rdf:type [ rdf:type owl:Restriction ; owl:onProperty spool:hasPTestStatus ; owl:someValuesFrom spool:Ongoing].}"""] .

Rule 2
# every spool has project status equal to value of pressure test status.
spool:Spool a owl:Class ;
 spin:rule [ a sp:Construct ;
 sp:text"""PREFIX spool: <http://www.spools.com/ontologies/spools.owl#>
    CONSTRUCT { ?this rdf:type [ rdf:type owl:Restriction ; owl:onProperty spool:trueXrayStatus ; owl:someValuesFrom ?XRay],
                                        [ rdf:type owl:Restriction ; owl:onProperty spool:truePTestStatus ; owl:someValuesFrom spool:NotStarted] ,
                                        [ rdf:type owl:Restriction ; owl:onProperty spool:trueProjectStatus ; owl:someValuesFrom spool:NotStarted].}
    WHERE {?this rdf:type [ rdf:type owl:Restriction ; owl:onProperty spool:hasXrayStatus ; owl:someValuesFrom ?XRay].
              ?x rdf:type [ rdf:type owl:Restriction ; owl:onProperty spool:hasPTestStatus ; owl:someValuesFrom spool:NotStarted].
              ?x owl:sameAs ?this}"""] .

Q4
# Selects and infers the correct spool statuses in reports
PREFIX spool: <http://www.spools.com/ontologies/spools.owl#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xml: <http://www.w3.org/XML/1998/namespace>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT ?spool ?XRay ?PTest ?Project
WHERE{?spool rdf:type [ rdf:type owl:Restriction ; owl:onProperty spool:truePTestStatus ;owl:someValuesFrom ?PTest].
?y rdf:type [ rdf:type owl:Restriction ; owl:onProperty spool:trueProjectStatus ;owl:someValuesFrom ?Project].
?y rdf:type [ rdf:type owl:Restriction ; owl:onProperty spool:trueXrayStatus ; owl:someValuesFrom ?XRay].
?spool owl:sameAs ?y
}
```

R4

spool	XRay	PTest	Project
16-FL-A12-0020001A	Completed	Ongoing	Ongoing
16-FL-A12-0020001B	Completed	Ongoing	Ongoing
18-P-A12-0020001A	Completed	Ongoing	Ongoing
18-P-A12-0020001C	Completed	Ongoing	Ongoing
18-P-A12-0020001B	Not Started	Not Started	Not Started
2-SV-A12-0020006	Completed	Ongoing	Ongoing

Figure 5.7: Rules and sample query with inference on integrated data to detect data quality issues

5.5 Design Evaluation

Acceptance criteria for evaluating the proposed framework are derived from the requirements elicited from the stakeholders (see Section 5.1). The simple assessment rule is that the solution should give positive indication on each criterion. The following assessments were carried out:

a) Time: Figure 5.8 compares the time expended in using the approach with the time taken to search through the reports manually and determine the true status of a spool. With a good understanding of all the necessary concepts and tools, it takes an hour to generate the case study project data, rules and queries. On the other hand, it takes about five minutes to look through all the reports and determine what the correct statuses of a spool should be. On projection, by the time this manual process is repeated 15 times, the upfront cost of developing the ontological solution would have been fully covered - because it takes about a minute to modify the query for any given spool and get results. Considering there are thousands of spools on large oil and gas fabrication projects according to Soleimanifar (2016), by the time the task on determining the statuses of a spool in the different reports is carried out 200 times (for example), the ontology-based solution would only have taken about 260 minutes compared to 1000 minutes in a manual process. This is an outright 70% saving, which will grow further as such task continues.

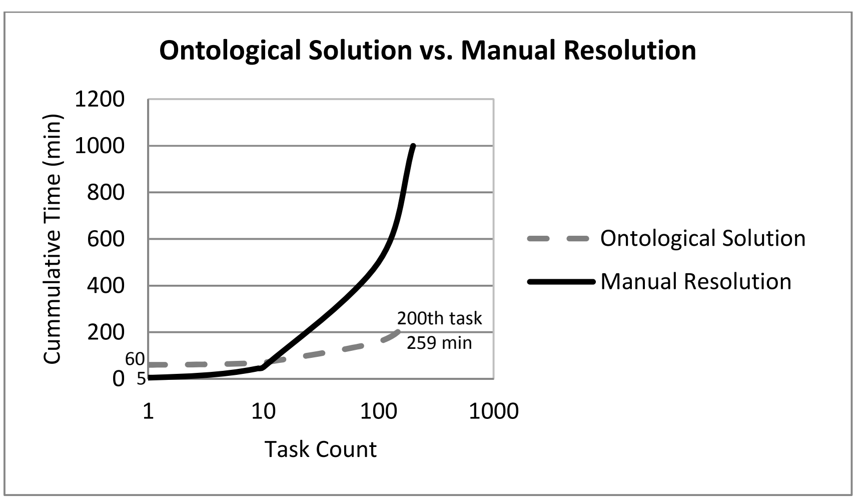


Figure 5.8: Cumulative duration comparison between ontological solution and manual process

b) Data quality checks: The use of SPIN rules in the proposed solution allows for automatic detection of data inconsistencies - something that is not possible in the manual inspection approach. In Figure 5.6, even though the data in the sources have been integrated and retrieved correctly, the output data is not logical. For example, it is not possible for spool 18-P-A12-0020001B to have a project status 'Ongoing' when the x-ray and pressure test statuses are 'Not Started'. This is so because the work steps requires that x-ray activities precede pressure testing and the pressure test report be used to populate the overall project report. By adding rules, as in the proposed method, inconsistencies are eliminated, and the correct result is outputted as in Figure 5.7.

c) Reuse and Extendibility: Typically, reports are updated periodically on a project. In such a scenario, the mappings, queries and rules will remain the same for the ontology-based framework. The updated data in different sources will only need to be reimported into the initial ontology. This eliminates the need for any reconciliation that may be required in the manual approach. Another advantage of the proposed framework is that people having similar work processes will not need to start from scratch. They may only need to adjust an existing template which will further reduce the setup cost. If the same template is used across multiple projects, the saving on the setup cost will further multiply. And when a template is well tested, it will become an organisation process asset over time.

5.6 Summary

This design showed that leveraging data integration technologies for process plant data management can be beneficial. By applying open source tools based on Semantic Web technologies, it is possible to make significant time savings compared with the traditional manual data management approach. The approach demonstrated is also low cost and responds to the need of small firms lacking the budget for proprietary technology tools. As demonstrated by the example, data in varying forms, originating from multiple sources, can be easily integrated; and the integrity of the output data will remain consistent with the original sources. Also, the mappings, rules, queries and designed ontology can be revised and reused multiple times. While there is an initial setup cost with the proposed approach, the prototype indicated such cost is quickly recovered within just a few tasks, and significant saving is achieved as the number of tasks increases. The next chapter evaluates two use cases based on the proposed solution.

Chapter 6

Evaluation Case Studies

From consultation with decommissioning industry stakeholders (Appendix 2) and reviewing of industry reports ((DNS, ZWS and ABB, 2015; DNS, AMEC and ZWS, 2016), two use cases standout. The first is with the prospect of reusing waste steel from offshore decommissioning in the construction industry. The other is with the prospect of piece small method of decommissioning which involves breaking the topside into small pieces offshore and transporting the waste to shore for processing. In the case of steel reuse, it is a more environmentally friendly approach to steel waste management compared to steel recycling. The related use case is on how to effectively assess the waste steel elements to determine their reuse potential. For the piece small removal method, the process allows the early removal of useful components from offshore installations to be decommissioned to limit the level of deterioration that could result from traditional decommissioning process, due to vessel wait times, and increase reuse opportunities. The related use case is on how to keep track of the piping and equipment components as they are being removed. Data integration proposals for these activities are here demonstrated and evaluated.

6.1 Piping and Equipment Audit

Strategising for decommissioning is not a one-size-fits-all scenario. Each facility is unique and will be considered separately for the appropriate decommissioning method. Consequently, decommissioning methods that could get more work done, safely and at a lower cost, are desired. This is because traditional methods, though get the work done, are very slow. One such new method is piece small removal technique (DNS, ZWS and ABB, 2015) that allows the early removal of useful components from installations to be decommissioned. The smaller the pieces of an offshore installation, the greater the options with regards to vessels and equipment needed for decommissioning. This eliminates vessel delay times. Getting items in good condition to shore early, after cessation of production, will limit the level of deterioration that could result from traditional decommissioning process.

In the piece small method, an important requirement is the recurring audit of the balance of items on the platform. The audit process involves the creation of an inventory of materials, components and equipment that will become available from the platform. The developed inventory details the quantity and qualities of these items to enable their tracking during decommissioning and identify those for re-use early in the process.

6.1.1 Data Sources

In this use case example, data on a process plant unit (Appendix 4) was sourced from an engineering service company that supports oil and gas asset life cycle information management (Ariosh, 2018). The data includes: maintenance record export from a maintenance database, information on operating parameters exported from the design model of the process plant unit, and piping isometric material take off (MTO) from the same design model.

6.1.2 Proposed Framework Application

The proposed data integration framework shown in Figure 5.1 is used for this use case. To integrate data, all the sources are pre-processed in spreadsheets and an ontology of all the available data created. Thereafter, available data from the sources are mapped to the appropriate entities in the ontology. This ensures a logical representation of the data sources is generated. The ontology, mapping file and integrated data for this task are provided in Akinyemi (2018c).

6.1.3 Use Case Evaluation

To test the proposed solution, the precision and recall method of benchmarking is used to measure its effectiveness in ad hoc query answering. The same querying tasks are also carried out, for comparison, using the recommended manual process (D3 Consulting Limited, 2017) for developing and managing waste inventory in the UK North Sea oil and gas industry. In this evaluation, five engineers that regularly carry out this type of exercise are tasked with using the industry recommended method. The first query in this evaluation seeks to determine the correct count of *600 pound Raised Face Weld Neck Schedule 40 2”* flanges on the different pipe lines in the integrated data. The proposed method answers this query using SPARQL. Query Q5 in Figure 6.1 is used for this and it simply says: select line number and quantity where (i) line number has a component with code 76.62.86.076.1 (the identifier for 600# RFWN SCH 40 2” flange), and (ii) selected line number has a component with a given quantity. The result for this query is shown under the semantic classifier column in the result count table in Figure 6.2. The same query exercise performed by the five engineers (referred to as classifiers) gives results as shown under columns labelled classifier1 to classifier5 in the result count table in Figure 6.2.

```
Q5
# Query to determine the correct count of all 600# RFWN SCH 40 2" Flanges

PREFIX : <http://www.plant.com/ontologies/plant.owl#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xml: <http://www.w3.org/XML/1998/namespace>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?lineNumber ?quantity
WHERE { ?lineNumber :hasComponentInfo ?component.
?component :hasItemCode :76.62.86.076.1.
?component :hasQuantity ?quantity
}
```

Figure 6.1: Query for task 1 on integrated data

Binary classification applies in this evaluation because there are only two possible predictions - positive and negative values. When a value is returned for a query, that indicates a positive prediction. Conversely when no value is returned, that suffices as a negative prediction. Using the result count values in Figure 6.2, the positive-negative predictions for all the classifiers are determined. To carry out a model-wide evaluation of the classifiers - that is, assess the performance of the classifiers at different thresholds, scores must be provided in addition to the predicted labels. The scores used in this evaluation are derived using variance calculation. Variance measures the spread of data (Erwin, 2011) and can be computed as shown in Equation 6.1. It is the sum of the squares of the terms (x^2) in a data set, divided by the number of terms in the set N, less the square of the mean (μ^2). A value of zero means no variability; a set of identical numbers have zero variance; small change means small variance; and large change means large variance. In order to score the predicted count values of classifiers 1-5, since the semantic classifier (proposed solution) has a perfect score –100% correct prediction, measures of the spread between counts returned by each of these classifiers and those of the semantic classifier are computed for each line number.

$$variance = \frac{\sum x^{2\cdot}}{N} - \mu^2 \qquad (6.1)$$

$$Score = 100 - variance \qquad (6.2)$$

Result Count

Line Number	Semantic Classifier	Classifier 1	Classifier 2	Classifier 3	Classifier 4	Classifier 5
30-P13003-61170X	7	2	7	7	7	7
30-P13001-61170X	7	2	7	7	7	7
30-P13002-61170X	20	2	20	20	20	20
30-P12019-61170X	2	2	2	2	2	2
8-P14547-61170X						
8-B62515-11170X	1	2	1	1	1	1
3-B62525-11170X	1	6	2	1	2	2
3-B62523-11170X	1	6	2	1	2	2
Total	39	22	41	39	41	41
Duration (min)	0	40	40	45	38	42

Positive-Negative Prediction

Line Number	Semantic Classifier	Classifier 1	Classifier 2	Classifier 3	Classifier 4	Classifier 5
30-P13003-61170X	P	P	P	P	P	P
30-P13001-61170X	P	P	P	P	P	P
30-P13002-61170X	P	P	P	P	P	P
30-P12019-61170X	P	P	P	P	P	P
8-P14547-61170X	N	N	N	N	N	N
8-B62515-11170X	P	P	P	P	P	P
3-B62525-11170X	P	P	P	P	P	P
3-B62523-11170X	P	P	P	P	P	P

Prediction Score

Line Number	Semantic Classifier	Classifier 1	Classifier 2	Classifier 3	Classifier 4	Classifier 5
30-P13003-61170X	100	93.75	100	100	100	100
30-P13001-61170X	100	93.75	100	100	100	100
30-P13002-61170X	100	19	100	100	100	100
30-P12019-61170X	100	100	100	100	100	100
8-P14547-61170X	100	100	100	100	100	100
8-B62515-11170X	100	99.75	100	100	100	100
3-B62525-11170X	100	93.75	99.75	100	99.75	99.75
3-B62523-11170X	100	93.75	99.75	100	99.75	99.75

Figure 6.2: Counts, predictions and scores outcome of the different classifiers on task 1

To relate the computed variance values to the perfect prediction scores obtained for the proposed solution, a score of 100 is used to indicate perfection, and computed variance values subtracted from 100 as shown in Equation 6.2. This transposes the variance values to scores that indicate higher values corresponding to higher probabilities toward positive count prediction and vice versa. The scores for all the classifiers are shown in Figure 6.2. Using the predicted labels and corresponding count scores, computation of the model-wide evaluation measures (precision, recall, F1 score and false positive rate) is carried out. Appendix 5 shows a sample precision and recall calculation while Appendix 6 gives detail on how they have been generated and plotted using a software package called Precrec (Saito and Rehmsmmeier, 2017).

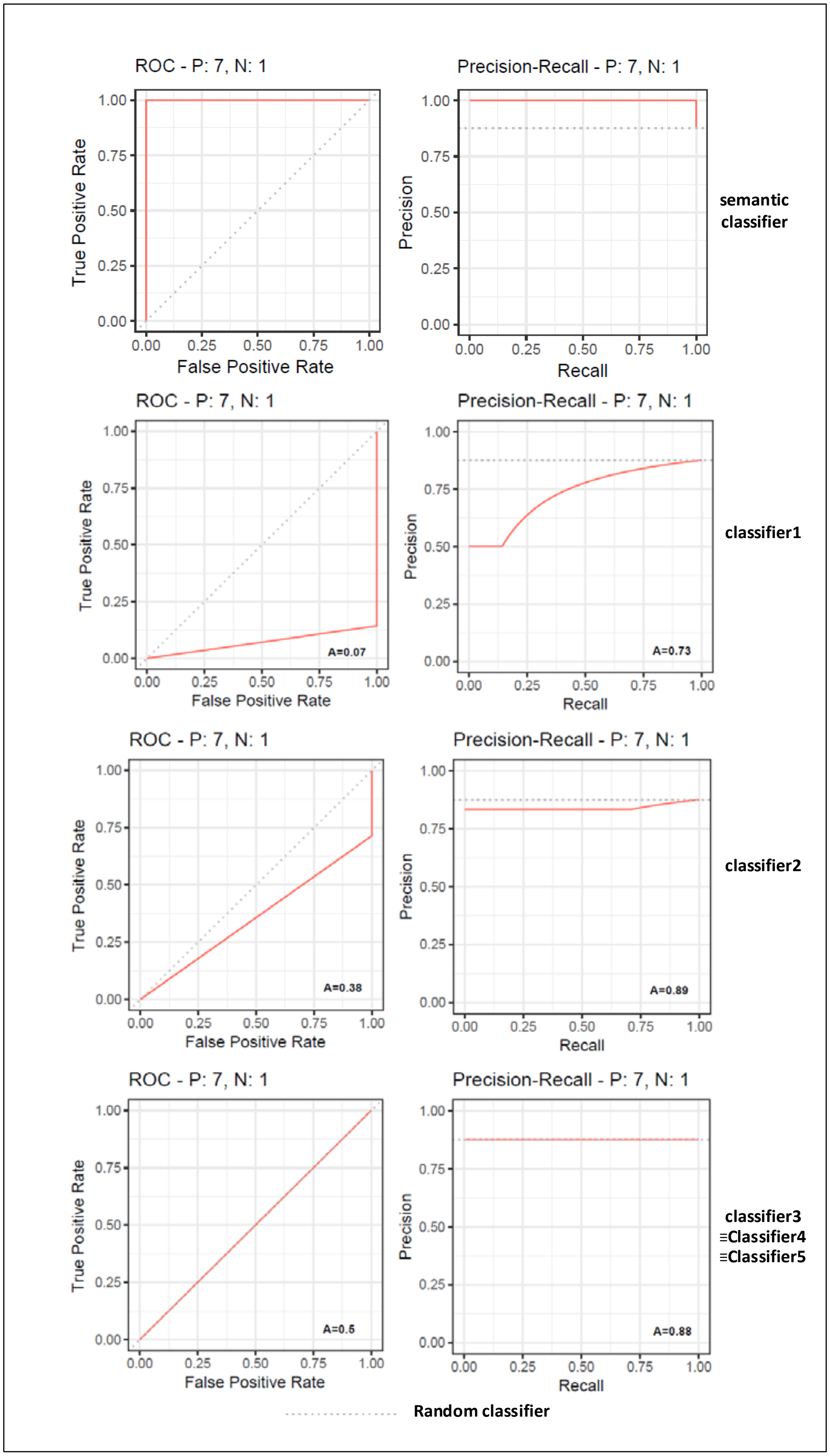


Figure 6.3: Precision and Recall outcomes of the tested classifiers on task1

Figure 6.3 shows the evaluation measures in graphs. Given that classifiers 4 and 5 share identical attributes with classifier 2, they have been left out. Also, given that the prediction data set is an imbalanced one (7 positives - P and 1 negative - N), a random classifier is located horizontally at (P/P+N) in all the precision-recall plots in Figure 6.3. This line divides the precision-recall space into two. The area above the line indicates good performance levels and the area below indicates poor performance. Similarly, the diagonal line starting at (0,0) and terminating at (1,1) is the random classifier for all the receiver operating characteristics (ROC) plots in Figure 6.3. It has the same classification function as the latter random classifier. Observation of all the plots in Figure 6.3 shows that classifiers 1 and 2 have poor performance, and classifier 3 is on the borderline. Given that the proposed method (semantic classifier) has a perfect performance, and the random classifier separates good and bad performances, then the proposed method outperforms the manual attempts. An alternative way to appraise the plots is to use the areas under the curves. By inspection, the total plot area of a ROC or precision-recall plot is 1. Given that the semantic classifier has a perfect performance, the area under any of its curves is equivalent to 1. This means that it clearly outperforms classifiers 1-5 considering the areas under the curves shown in Figure 6.3.

One way of getting the optimal blend of precision and recall for task 1 is to combine the two metrics using F1 score. At threshold 100 (normalised rank = 0), the least number of positive label instances occurs, and the values of recall and precision are closer to zero. As the threshold decreases, recall increases and more positive labels are discovered. Generally, in task 1, precision increases with recall because the number of false positives stays the same as the number of true positives increases. Consequently, at threshold 0 (normalised rank = 1), recall is total, and precision is highest. Given that the F1 score combines these metrics, it is highest for classifiers 1,2 and 3 at normalised rank of 1 as shown in Figure 6.4. However, these F1 scores are less than 1, which is expected in a perfect classifier. Consequently, the proposed method (semantic classifier) is better than the manual attempts by the engineers. See Table A.3 in Appendix 5 for context of this analogy.

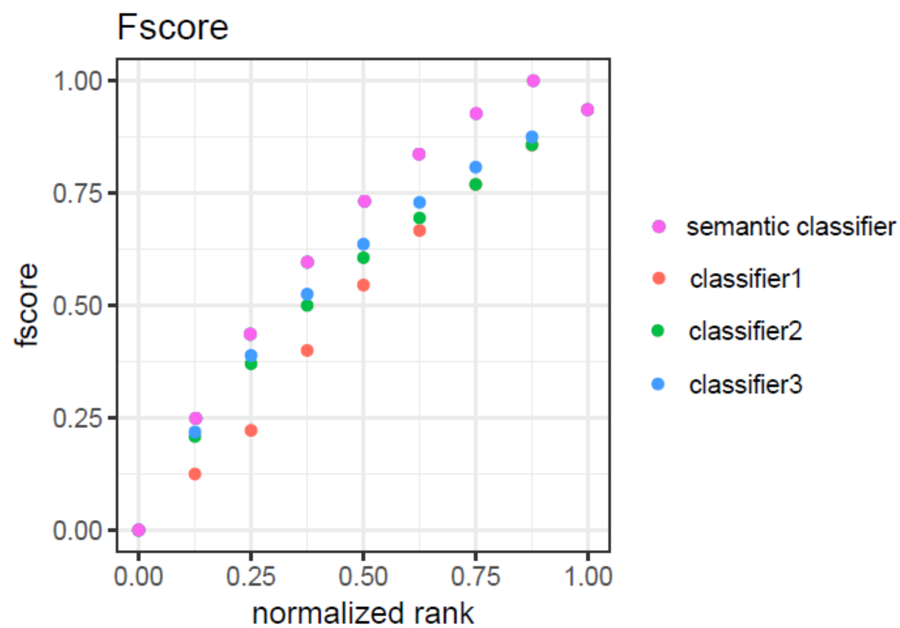


Figure 6.4: F1 score plot of tested classifiers on task 1

To validate the above results, the above test is repeated for another query. The second query in this evaluation seeks to determine the correct count of the same *600 pound Raised Face Weld Neck Schedule 40 2" flanges* on the different pipe lines but managed by 'Gbenga'. Query Q6 in Figure 6.5 answers this by simply saying: select line number and quantity where (i) staff ID AR010 is the ID for the person 'Gbenga', (ii) a given equipment is the responsibility of person with staff ID AR010, (iii) same equipment has a unique facility number, (iv) a line number is within the domain of a facility with number as previously listed, (V) the latter line has a component, (vi) the component has item code 76.62.86.076.1 (the identifier for 600# RFWN SCH 40 2" flange), and (vii) the selected component has quantity attribute.

```
Q6
#Query to determine the correct count of all 600# RFWN SCH 40 2" Flanges managed by Gbenga

PREFIX : <http://www.plant.com/ontologies/plant.owl#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xml: <http://www.w3.org/XML/1998/namespace>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT DISTINCT ?lineNumber ?quantity
WHERE {:AR010 :hasFirstName "Gbenga"^^xsd:string.
?equipment :hasResponsiblePerson :AR010.
?equipment :hasFacilityNo ?facility.
?lineNumber :hasFacilityNo ?facility.
?lineNumber :hasComponentInfo ?component.
?component :hasItemCode :76.62.86.076.1.
?component :hasQuantity ?quantity
}
```

Figure 6.5: Query for task 2 on integrated data

The result for this query is as shown under the semantic classifier column in the result count table in Figure 6.6. The same query exercise is also performed by the same five engineers (referred to as classifiers), and the results are as shown under columns labelled classifier1 to classifier5 in the result count table in Figure 6.6. Binary classification as in the previous task is also carried out and prediction scores computed from Equations 6.1 and 6.2 outputted prediction labels and scores are as shown in Figure 6.6.

Result Count

Line Number	Semantic Classifier	Classifier 1	Classifier 2	Classifier 3	Classifier 4	Classifier 5
30-P13003-61170X	7	2	7	7	7	7
30-P13001-61170X	7	2	7	7	7	7
30-P13002-61170X	20	2	20	20	20	20
30-P12019-61170X						
8-P14547-61170X						
8-B62515-11170X						
3-B62525-11170X						
3-B62523-11170X						
Total	34	6	34	34	34	34
Duration (min)	0	62	57	65	50	60

Positive-Negative Prediction

Line Number	Semantic Classifier	Classifier 1	Classifier 2	Classifier 3	Classifier 4	Classifier 5
30-P13003-61170X	P	P	P	P	P	P
30-P13001-61170X	P	P	P	P	P	P
30-P13002-61170X	P	P	P	P	P	P
30-P12019-61170X	N	N	N	N	N	N
8-P14547-61170X	N	N	N	N	N	N
8-B62515-11170X	N	N	N	N	N	N
3-B62525-11170X	N	N	N	N	N	N
3-B62523-11170X	N	N	N	N	N	N

Prediction Score

Line Number	Semantic Classifier	Classifier 1	Classifier 2	Classifier 3	Classifier 4	Classifier 5
30-P13003-61170X	100	93.75	100	100	100	100
30-P13001-61170X	100	93.75	100	100	100	100
30-P13002-61170X	100	19	100	100	100	100
30-P12019-61170X	100	100	100	100	100	100
8-P14547-61170X	100	100	100	100	100	100
8-B62515-11170X	100	100	100	100	100	100
3-B62525-11170X	100	100	100	100	100	100
3-B62523-11170X	100	100	100	100	100	100

Figure 6.6: Counts, predictions and scores outcome of the different classifiers on task 2

Figure 6.7 shows the evaluation measures, of the classifiers for the second query, in graphs. Given that classifiers 2-5 have identical outcomes, they are represented by classifier 2. Also, given that the prediction data set is imbalanced (3 positives - P and 5 negatives - N), the random classifier is located horizontally at (P/P+N) in the precision-recall plots in Figure 6.7. Similarly, the diagonal line starting at (0,0) and terminating at (1,1) is the random classifier for the ROC plots in Figure 6.7. Inspection of the plots shows that classifier 1 has poor performance, because it occupies the space below the random classifier; and classifier 2 is at par with the random classifier. However, given that the proposed method (semantic classifier) has a perfect performance, then it clearly outperforms the manual attempts. An alternative way to appraise the plots is to use the areas under the curves. Considering that the semantic classifier has a perfect performance, the area under any of its curves is equivalent to 1. This means that it clearly outperforms classifiers 1-5 considering the areas under the curves shown in Figure 6.7.

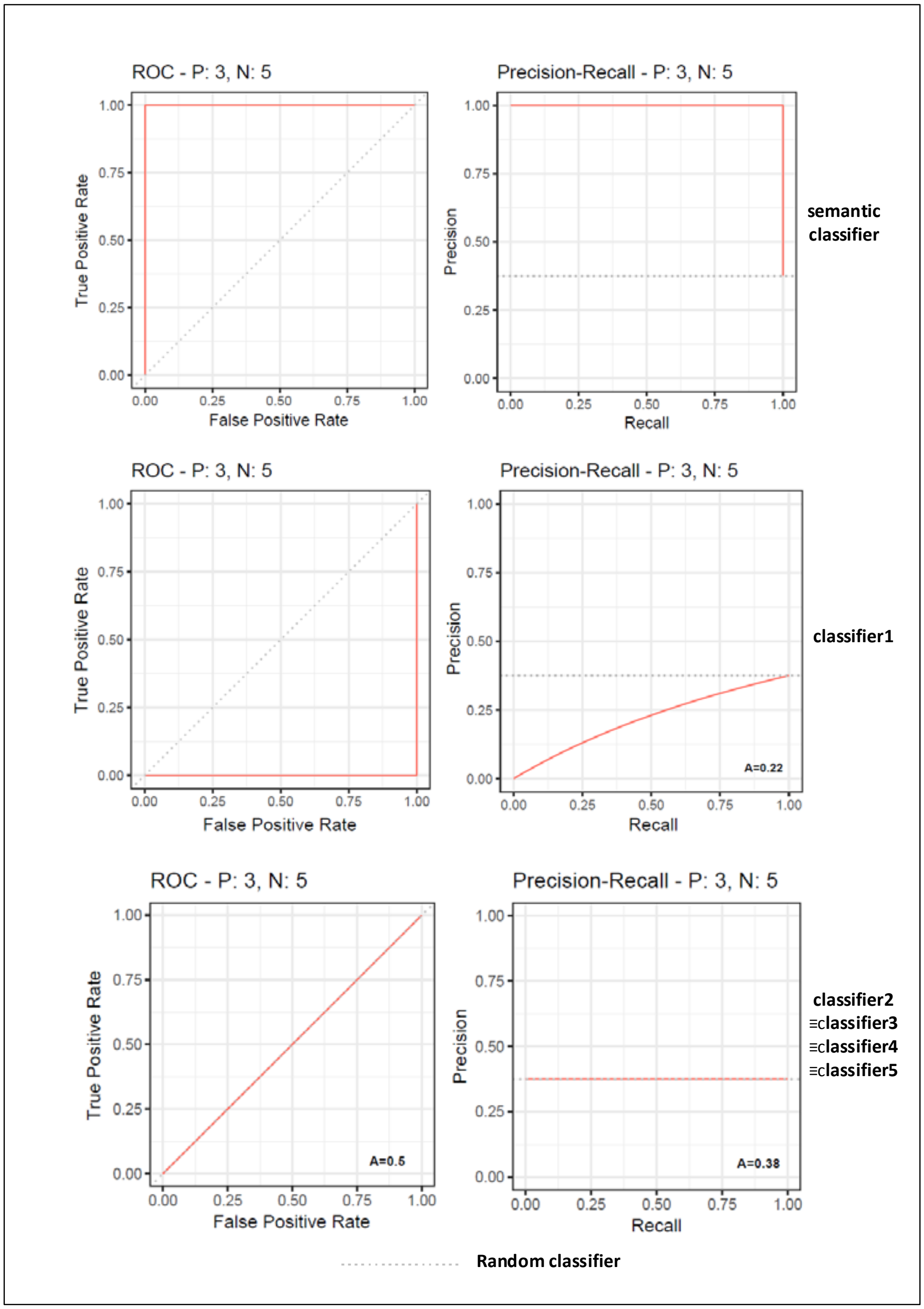


Figure 6.7: Precision and Recall outcomes of the tested classifiers on task2

To get the optimal blend of precision and recall in task 2, the F1 score is used as a measure. For classifier 1 on task 2, precision increases with increase in recall (see Figure 6.7) as the classification threshold decreases. However, the highest F1 score observable in Figure 6.8 for classifier 1 is less than 1, which is expected in a perfect classifier. For classifier 2 on task 2, precision remains the same (see Figure 6.7) as recall and classification threshold vary. Again, the highest F1 score observable in Figure 6.8 for classifier 2 is less than 1, which is expected in a perfect classifier. Consequently, the proposed method (semantic classifier) is better than the manual attempts by the engineers. See Table A.4 in Appendix 5 for context of this analogy.

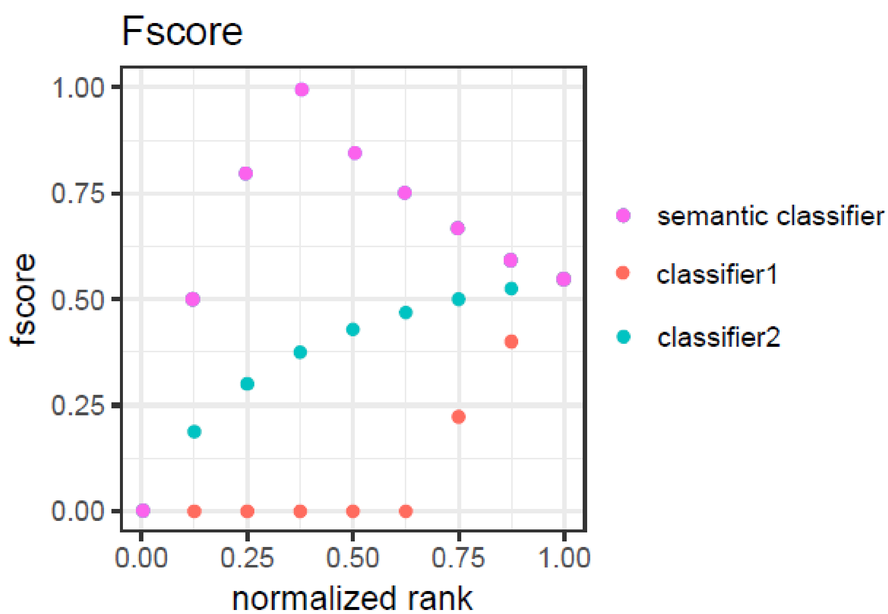


Figure 6.8: F1 score plot of tested classifiers on task 2

Finally, a comparison of the total duration expended by each engineer and the proposed solution is shown in Figure 6.9. While accurate results can be returned within 40 mins for the two tasks using the proposed semantic classifier, it takes about 90 mins for an engineer to get results, that are sometimes not completely correct, for similar tasks.

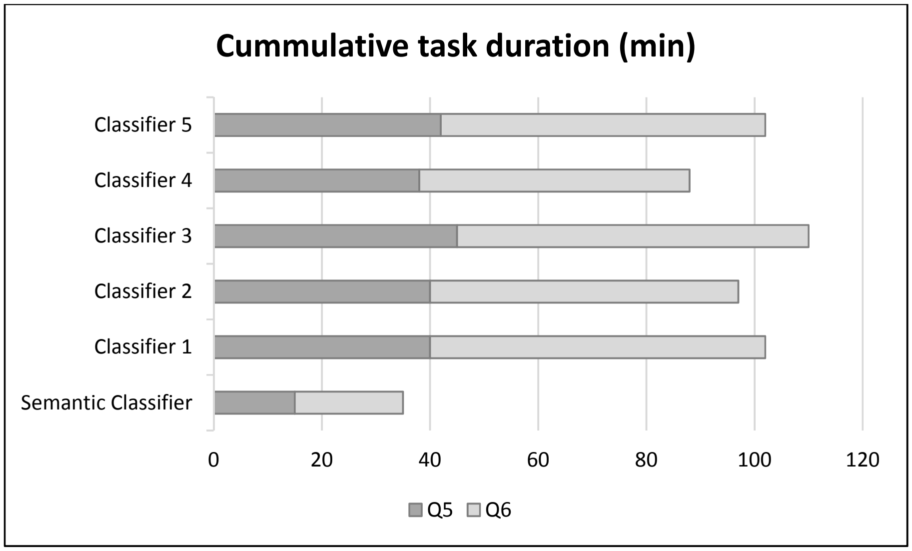


Figure 6.9: Cumulative duration comparison between tested classifiers

The results of this evaluation indicate that the current approach of managing inventories of decommissioned materials is inefficient, considering the time taken to complete tasks. Also, considering that none of the engineers (classifiers 1-5) achieved an F1 score of 1, outcomes from tasks using the recommended spreadsheet are error-prone, irrespective of the thoroughness of the user. The result of this evaluation is limited by the size and complexity of the data sample. On a full-sized data set e.g. complete topside of an offshore oil and gas production platform, classifiers 1-5 have a very limited chance of successfully managing the data. On the other hand, the semantic classifier has a good chance of success. However, this would depend on the robustness of the semantic application itself (Gray *et al.*, 2014).

6.2 Structural Steel Assessment

The Royal Society for the encouragement of Arts, Manufactures and Commerce (RSA) et. al (2015) reported that a typical North Sea platform is about 97% steel and an estimated 500,000 tonnes of the over 5 million tonnes of steel - including platform topsides, jackets and subsea components - presently installed, will be decommissioned before 2023 (RSA and ZWS, 2015). These assertions indicate that steel is going to be a large portion of the waste from offshore decommissioning. Decommissioning projects often report up to 98% recycle and reuse of waste materials. However, only about 2-3% are reused and the benefits of recycling are significantly reduced because of the energy consumed in the process (DNS, AMEC and ZWS, 2016). A survey in 2008 discovered that 0.2% of 9 million tonnes of steel scrapped was reused (WellMet2050, 2010). This indicates that steel reuse is very limited. However, an earlier survey in 1998 found that the amount of steel reused in the construction sector was 31% of 129,000 tonnes sent for recycling (Bioregional, 2008). This underscores the potential reuse opportunity of structural steel from decommissioned offshore facilities. As steel jackets and subsea components of offshore facilities may be disposed at sea, depending on the scenario (DECC, 2011), this research has focused on the topside structural steel (example shown in Figure 6.10) because there is clarity on their complete removal from sea (ARUP, Scottish Enterprise and DNS, 2014).

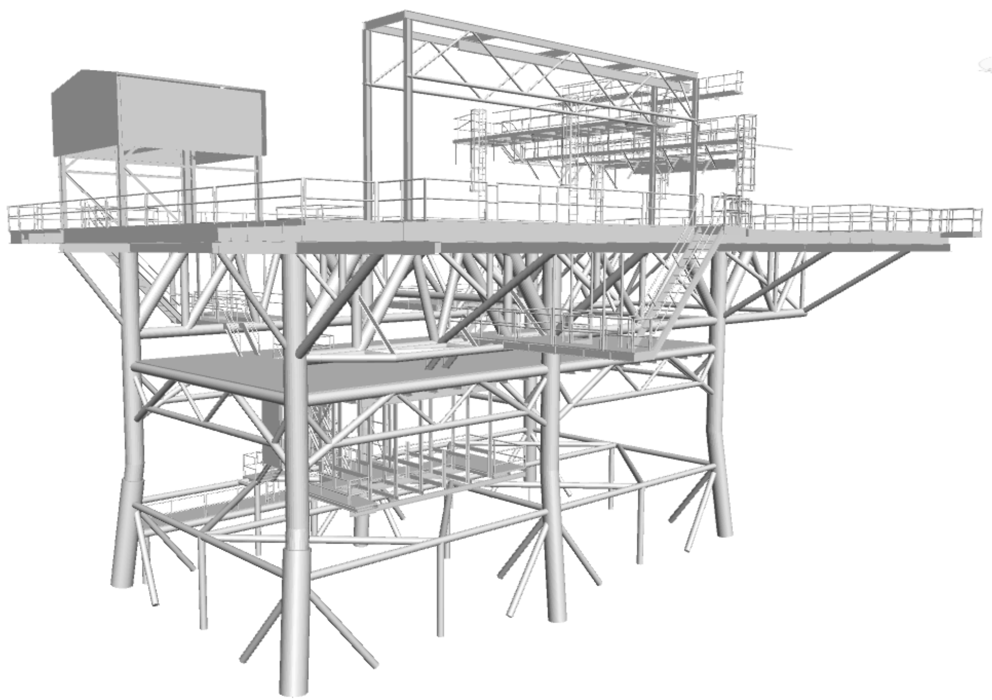


Figure 6.10: Steel carcass of the topside of an offshore platform

For decommissioned steel sections to have application in construction, they need to have traceability and documentation (DNS, AMEC and ZWS, 2016). Assessment information specified by specialist asset re-sale companies include certifications, warranties, material inventory, fabrication history record, condition report, 2D drawings and 3D models (DNS, AMEC and ZWS, 2016). These also need to be corroborated with virtual 3D images from 3D scanning (DNS, ZWS and ABB, 2015). Information obtained from the sources is used to determine the quality, condition, age and coating of fabricated steel. However, considering the heterogeneity of the sources, building an asset inventory to efficiently deliver the required assessment is always a challenge (DNS, ZWS and ABB, 2015; 2016). This evaluation compares the state-of-the-art method (see Section 3.3) of carrying out the required assessment to a solution for handling it based on the proposed framework.

6.2.1 Data Sources

This case study makes use of synthetic data generated based on a real offshore platform documented by Ariosh (2018). The data sources are three of the commonly used sources for steel assessment. In Figure 6.11, Source 1 (S1) is a structural report from a 3D application; source 2 (S2) is a maintenance database on the steel sections in S1; and source 3 (S3) is a summary of 3D point cloud data scan positions. S1 contains the description, specification, geometric properties and location of reported steel members. S2 contains the details of maintenance service, the contact of service provider and specific remark for instances of maintenance activity carried out. S3 lists the primary scan positions links that are closest to the zones of a platform to which a steel member belongs. From these scan positions, an assessor can see related scan positions and very quickly peruse the zone of interest. This case study has limited instances in the sources for demonstration purpose. For context, the items listed in the report have an approximate weight of 113 tonnes compared to single North Sea platforms that weigh up to 48,000 tonnes (DNS, ZWS and ABB, 2015) and over. From the data sources, there is a need to be able to deduce all the relevant information required for assessment of a steel section for reuse.

S1: Structural report from facility model

ID	Description	Spec	Cut Length	Area	Zone
B.01	Circular Hollow Section	CHS 168.3x5.6	14 m		Boat Deck
B.02	Circular Hollow Section	CHS 168.3x5.6	14 m		Boat Deck
B.03	Circular Hollow Section	CHS 219.1x5.0	18 m		Boat Deck
B.04	Circular Hollow Section	CHS 219.1x5.0	17 m		Boat Deck
B.05	Circular Hollow Section	CHS 323.9x5.6	18 m		Boat Deck
B.06	Circular Hollow Section	CHS 323.9x5.6	17 m		Boat Deck
B.07	Circular Hollow Section	CHS 355.69x8.0	20 m		Boat Deck
S.01	Universal Beam	UB 762x267x226	10 m		Boat Deck
S.02	Parrallel Flange Channel	PFC 260x75x28	15 m		Boat Deck
S.03	Grating	38x5 SG		98 sq m	Boat Deck
C.01	Circular Hollow Section	CHS 168.3x5.6	14 m		Cellar Deck
C.02	Circular Hollow Section	CHS 219.1x5.0	18 m		Cellar Deck
C.03	Circular Hollow Section	CHS 323.9x5.6	18 m		Cellar Deck
C.04	Circular Hollow Section	CHS 355.69x8.0	20 m		Cellar Deck
C.05	Universal Beam	UB 762x267x226	15 m		Cellar Deck
C.06	Parrallel Flange Channel	PFC 260x75x28	14 m		Cellar Deck
C.08	Grating	38x5 SG		150 sq m	Cellar Deck
M.01	Universal Beam	UB 762x267x226	15 m		Main Deck
M.02	Parrallel Flange Channel	PFC 260x75x28	15 m		Main Deck
M.03	Grating	38x5 SG		70 sq m	Main Deck

S2: Structural maintenance database

maintenance

id	structure	structure Id	serviceDate	serviceId	remarks
M001	Boat Deck	X.BD	9/5/2010	S003	Repairs done
M002	Grating	S.03	7/5/2011	S001	Replaced
M003	Cellar Deck	X.CD	7/6/2013	S001	Repairs done
M004	Main Deck	X.MD	7/27/2013	S001	Repairs done
M005	Boat Deck	B.07	11/15/2015	S002	Fatigued
M006	Grating	S.03	9/18/2017	S002	Corroded and needs replacement

contact

id	company	phone	website
C0001	UTEC	+441224 812020	http://utecsurvey.com
C0002	Siem	+44 1224 945500	https://siemoffshorecontractors.com/

services

id	contactId	label
S001	C0001	Topside inspection and maintenance
S002	C0001	Structural inspections and cleaning
S003	C0002	Splash zone inspection and maintenance

S3: 3D scan summary of facility laser model

Platform Level	Virtual 3D
Boat Deck	C:\Navisworks Data\Delta PP\Revised Truview with Model Superimposed\DATA\BOAT & SUBCELLAR DECK.htm
Cellar	C:\Navisworks Data\Delta PP\Revised Truview with Model Superimposed\DATA\CELLAR DECK.htm
Main Deck	C:\Navisworks Data\Delta PP\Revised Truview with Model Superimposed\DATA\MAIN & UPPER DECK.htm

Figure 6.11: Sample facility steel data on design specification, maintenance and 3D virtual image

6.2.2 Proposed Framework Application

The proposed data integration framework for decommissioning waste management makes use of ISO 15926 because the standard is focused on integrating life cycle data of oil and gas facilities (Fiatech, 2011). The typical data sources for a process plant include data file exports from software applications, asset databases and project websites (Yang, 2009). A scenario where access to asset databases is granted is also possible (Fiatech, 2011). These sources are usually in formats other than RDF which is the recommendation for ISO 15926. As a result, this proposal includes recommendations for handling data conversion. A very important advantage of ISO 15926 is that, being a global standard, most stakeholders and application developers will seek to comply with it. However, the difficulty with using the standard, for most engineers, is identifying how to couple the available tools with the requisite technologies for its proper use (Fiatech, PCA and USPI, 2015). Also, the lack of examples on available deployments may not encourage potential users.

Figure 6.12, depicts the proposed data integration framework for oil and gas asset data. The framework uses ISO 15926 as the common data model for the sources. To integrate data, information patterns made from generic templates based on Parts 2, 4, 7 and 8 of ISO 15926 standard are used to represent the ontology of the sources. Thereafter, data from the sources are mapped to the ontology to generate RDF data. The RDF data is then deposited in a data repository where users and applications can interface with it via SPARQL queries, the standard query language for RDF data. These queries, which answer questions relevant to the purpose for which the framework is deployed, can be documented as result templates that are reusable. The generic steps of the work process for the proposed framework are described below:

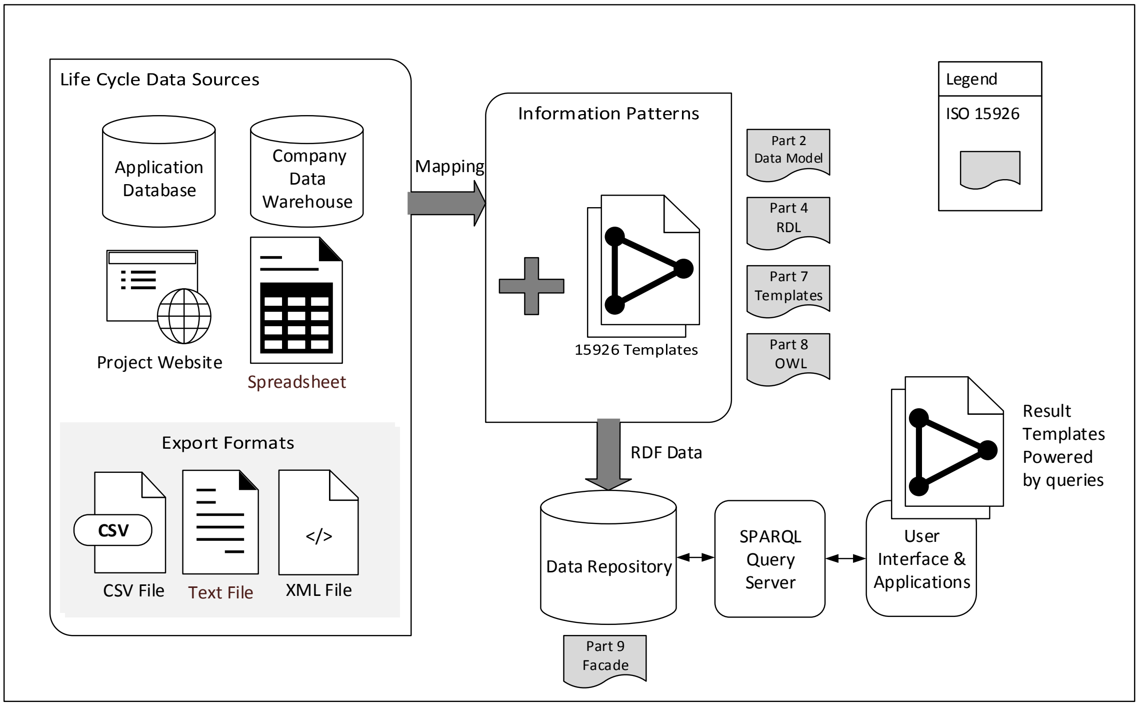


Figure 6.12: Data integration using ISO 15926 ontology

a) *Pattern Definition***:** This is in part identifying the right templates from those already created (Parts 7 and 8) or formulating a new one (Parts 2 and 4) to represent the entities in the sources. The other part is generating information patterns that represent the ontology of the data sources. Information patterns can be created using tools like dot15926Editor (TechInvestLab.ru, 2013) and Protégé (Musen, 2015).

b) *Data mapping*: This involves organising the sources and generating the mapping files for moving the data in the sources into the information patterns. Mapping of data from the sources can be done using tools like Cellfie (Protégé Project, 2017) and Ontop (Calvanese *et al.*, 2017). Mapping of data can also be done using open-standard file format like JavaScript Object Notation (JSON) (TechInvestLab.ru, 2013) or a mapping language like 'Relational Database to RDF Mapping Language' (R2RML) (Das, Sundara and Cyganiak, 2012 W3C Recommendation).

c) *Data Storage*: This is about the architectural decision on where to locate the prepared data for the querying process. The sources could include files located on the local machine, triple stores or relational databases. A decision must be made on whether to keep all the source together (warehousing) or link them up from different administrative entities (virtualisation). ISO 15926 Part 9 is supposed to address data storage using triple stores but is yet to be released. Nonetheless, the idea around using triple stores for storing RDF data has become well established (Knap *et al.*, 2018). The data format for the proposed framework is RDF. As a result, accessing a relational database requires an Ontology-Based Data Access (OBDA) model (Calvanese *et al.*, 2009). This helps to relate the data in the relational database to the appropriate ontology entities.

d) *Information Extraction*: SPARQL is used for information extraction because data is in RDF format. When prepared data is in a single location, normal SPARQL queries can be used for information extraction. However, when querying across multiple sources, SPARQL federated queries (Prud'hommeaux *et al.*, 2013 W3C Recommendation) are used. Where relational databases are maintained as a source, the OBDA approach can be used to on-the-fly translate the relevant part of the SPARQL query into a local query. Ontop is one such tool (Calvanese *et al.*, 2017).

6.2.3 Proposed Solution

To realise the proposed data integration framework for decommissioning waste management, open source applications are coupled for demonstration purposes. Figure 6.13 shows the plan for the data sources. S1 and S3 are exports that are available as data files while S2 is assumed to be residing in an active relational database. For the sources that are data files, the data in the sources is mapped to information patterns based on ISO 15926 templates and standard entities. The resulting RDF data is then deposited in a triple store. For the source that is a relational database, the source is sustained as-is; and SQL queries are used to extract the data of interest. The data from the relational database is made available as RDF data in a triple store using an OBDA model which links the data to appropriate ontology entities. SPARQL queries are then used to extract information of interest from the repositories. Several permutations of these RDF data generation methods are possible, and any combination of data files, relational databases and triple stores as sources will work. The two important steps include (i) generation of the ISO 15925 information pattern for the data in the sources and (ii) generation of mapping files to migrate the data in the sources into the information patterns for RDF data generation.

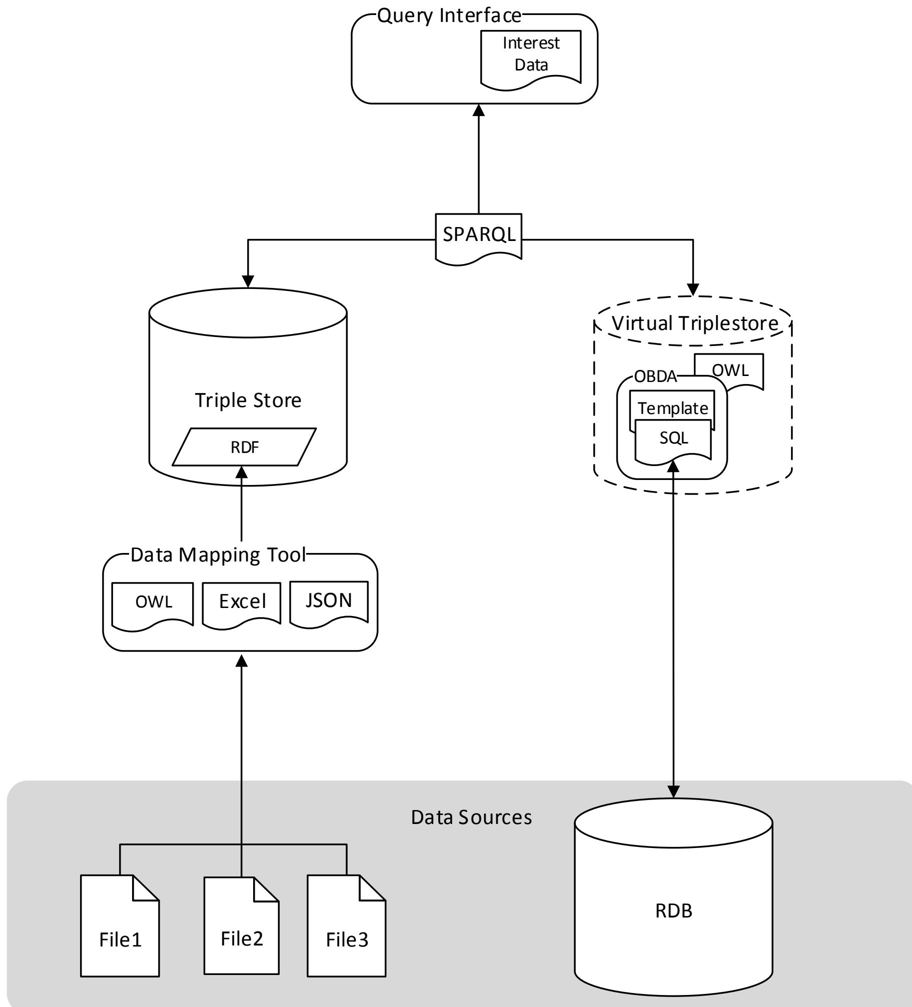


Figure 6.13: Implementation architecture for case study

The open source tools used in the implementation include dot15926 Editor (TechInvestLab.ru, 2013), RDF4J (Knap *et al.*, 2018), Ontop (Calvanese *et al.*, 2017) and Protégé (Musen, 2015). dot15926 is an architecture and set of specific libraries that enable working with ISO 15926. It facilitates reading, visualizing, exploring, searching, reasoning, mapping etc. of ISO 15926 data (TechInvestLab.ru, 2013). It was used for RDF data generation for the data file sources in this case study. RDF4J is a Java framework for storing, reasoning, querying etc. RDF data (Knap *et al.*, 2018). It was used as the RDF data repository and SPARQL query server. Ontop is an OBDA system that allows querying relational data sources provided in terms of an ontology to which the data sources are mapped (Calvanese *et al.*, 2017) and Protégé is an ontology editor that supports the latest OWL standard (Musen, 2015). The latter two tools work together. Ontop is a plug in protégé and the two tools help in extracting data from the relational database and generating RDF data in a virtual triple store.

Data Processing

To map data from the sources, mappings files and information patterns must be developed. With the use of the dot15926 Editor, the data file type sources, preprocessed into spreadsheets, are converted to RDF data. The first step in this process, is to identify the entities in the sources. By inspection of the sources S1 and S3 in Figure 6.11, eleven entities including 6 types of steel sections, a grating type, 3 deck types and point cloud representation must be identified in ISO 15926 Reference Data Library (RDL) (TechInvestLab.ru, 2013). Identifying them in the RDL will ensure that anyone that deals with this data can reference the same meaning and capture the exact description. The eleven entities identified are the reference data for the sources. The spreadsheet with header H1 shown in Figure 6.14 is prepared for the project reference data. The grey cells in the header are added as columns to the base information derived from the sources. A namespace (qn1) for the resources is concatenated with the Local IDs generated for them to form URIs[14]. This is compliant with best practices for identifiers (McMurry *et al.*, 2017). Also, the Part 2 type of the entities, their Part 4 RDL superclasses and the superclasses' URIs are logged. The qualified name for the Part 4 superclasses is pcardl. This is the standard namespace for the RDL maintained by POS Caesar Association (PCA), an organisation that maintains the integrity of the registered Part 4 entities (Fiatech, 2011; TechInvestLab.ru, 2013).

[14] URI ≡ IRI. URI works on ASCII characters while IRI works using a Unicode typeset.

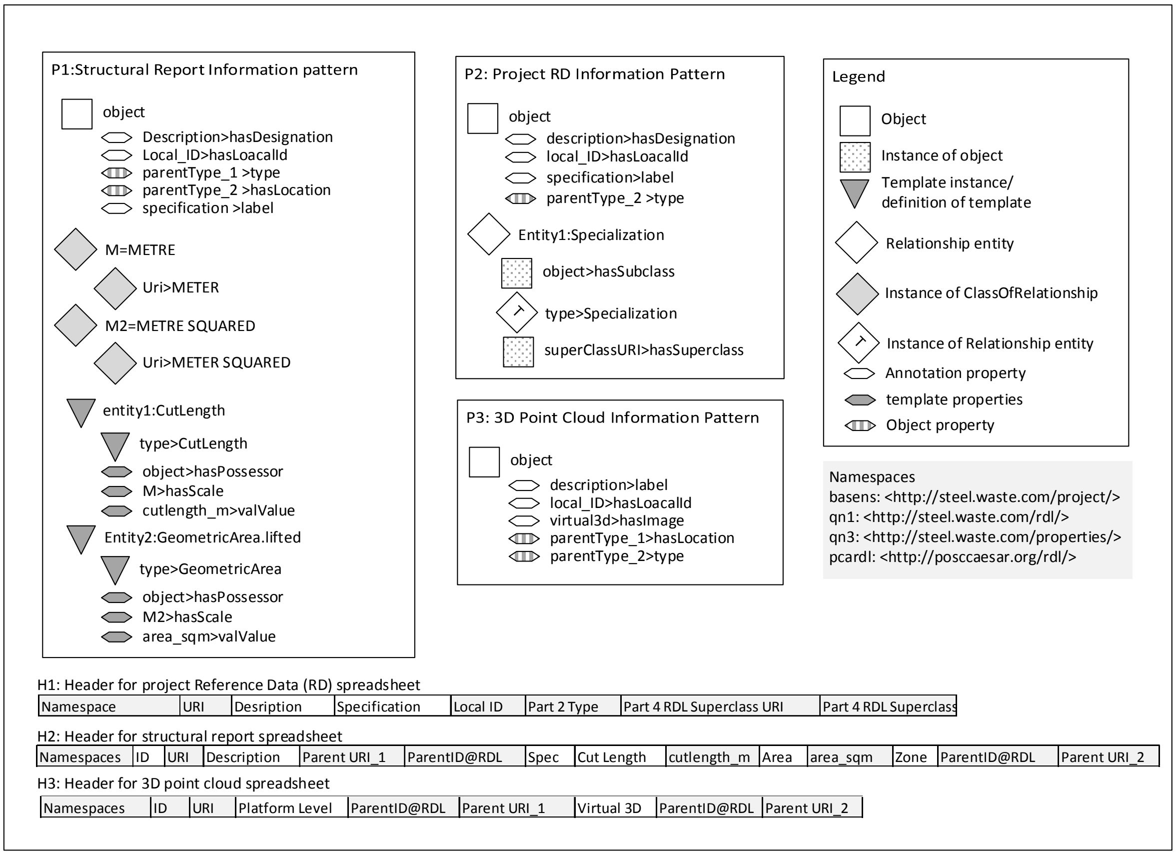


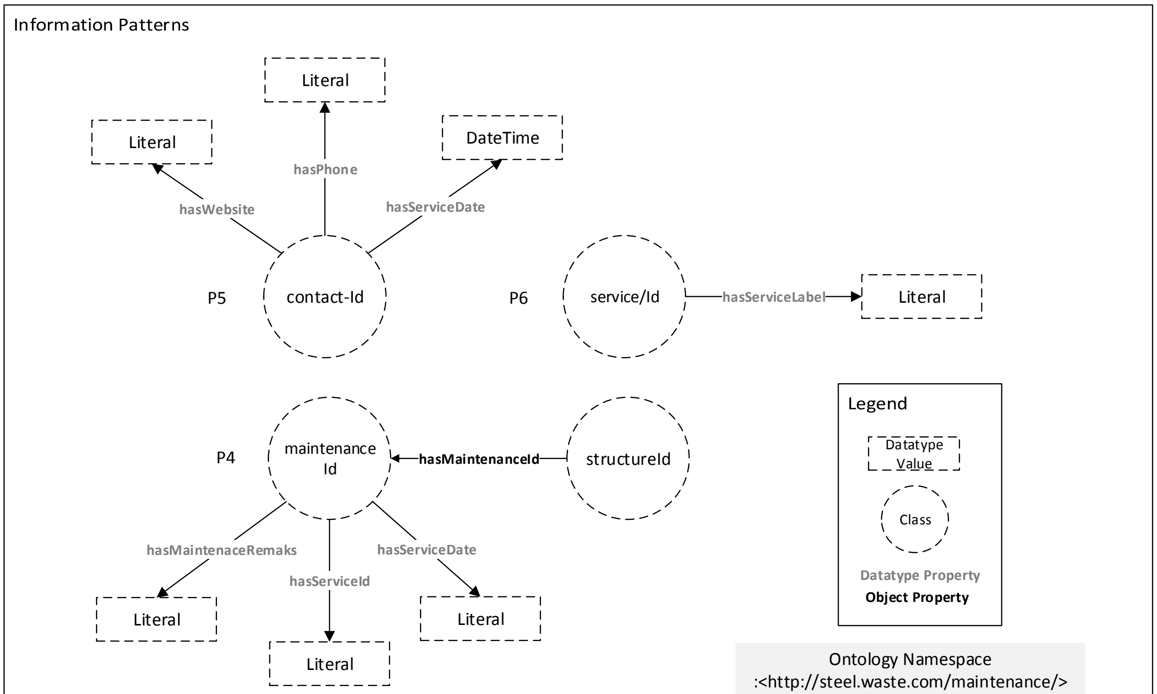
Figure 6.14: ISO 15926 information patterns for structural report and 3D scans

For data in S1, the spreadsheet with header H2 as shown in Figure 6.14 is prepared. Namespace (qn3) is concatenated with the ID of the resources to generate URIs for them. For each resource in the source, the project reference data entities and their corresponding URIs are regenerated in new columns. Lastly, the values of cut length and area for the resources are extracted into separate columns. For data in Source S3, the spreadsheet with header H3 is prepared. Again, namespace (qn3) is concatenated with the ID generated for the resources to create URIs for them. Also, the project reference data entities and their corresponding URIs are regenerated in new columns. Once the data sources are prepared, the information patterns are developed based on them. For S1, the pattern P1 in Figure 6.14 is created. It has a signature where its object URI has annotation properties, ‘hasDesignation’, ‘hasLocalId’ and ‘label’. Also, the same object URI has object property ‘type’ and ‘hasLocation’. Two Part 7 templates are also in the information pattern. There is a template for cut length and another for geometric area. These are to capture the records relating to length and area of entities. Each template has properties for the possessor of the record, the record's scale of measurement and the value of the measurement. Also, the ISO 15926 Part 4 representation of the measurement scales are instantiated in the pattern.

For source S3, the information pattern P3 in Figure 6.14 is created. It has a signature where its object (URI) has annotation properties, 'label', 'hasLocalId' and 'hasImage'. The same object has object property, 'type' and 'hasLocation'. For project reference data entities, an information pattern P2 is created. It has a signature where its object has annotation properties, 'hasDesignation', 'label' and 'hasLocalId'. In addition, a Part 2 specialization entity showing the relationship between the resources and their Part 4 superclasses is added to the pattern. Once the patterns are ready, mapping files are generated to map the data from the sources into the patterns. JSON was used for representing the mappings in the case study and RDF data was generated from the process. Appendix 7 details the mapping process from sources S1 and S3.

Regarding the mapping of data from the relational database S2, an OBDA model and an ontology of the entities are required. Guidance for implementing OBDA is provided in Appendix 8. The OBDA model contains mapping axioms comprising SQL queries and target triple templates. The mapping file can either be generated using R2RML and imported into Ontop or automatically generated by Ontop itself using its native mapping language (Calvanese *et al.*, 2017). The triple templates of the mapping axioms in this case study are contained in information patterns P4, P5 and P6 generated for the maintenance database and shown in Figure 6.15. P4 has a signature with an object property 'hasMaintenanceId' linking objects that belong to classes 'structureId' and 'maintenaceId'. Objects that belong to the latter class in addition have datatype property 'hasMaintenanceRemarks', 'hasServiceId' and 'hasServiceDate'. Pattern P5 has a signature where objects in class 'contact-Id' have datatype property 'hasWebsite', 'hasPhone' and 'hasServiceDate'. The last information pattern P6 has a signature where objects that belong to the class 'service/Id' have datatype property 'hasServiceLabel'.

The complementing part of the mapping axioms used for extracting data from source S2 are the corresponding SQL queries for mappings M1, M2, M3 and M4 shown in Figure 6.15. The SQL query for M1 selects values of 'id' and 'structureId' from the maintenance table in S2. The SQL query for M2 selects values of 'id', 'serviceId', 'serviceDate' and 'remarks' from the same table. The SQL query for M3 selects 'id', 'company', 'phone' and 'website' from the contact table in S2; and the SQL query for M4 selects 'id' and 'label' from the services table in S2. All the resources from source S2 use the ontology namespace <http://steel.waste.com/maintenance/> except for objects of class 'structureId' in M1 that have a namespace (qn3) that is the same with the resources from sources S1 and S3. It should be noted that resources that share the same URI in a project are one and the same. This is a consequence of representing the integration data model in RDF. The generated files from this process are deposited in a custom Ontop triple store in the RDF4J framework. This triple store behaves like a virtual one because the actual RDF data does not reside there. Instead, it is generated on the fly when the database query is executed.



Mapping	Target TripleTemplate	Source SQL Query
M1	<http://steel.waste.com/project/{structureId}> **:hasMaintenanceId** :{Id} .	SELECT "id", "structureId" FROM "maintenance"
M2	:{Id} :hasServiceDate {serviceDate} ; :hasServiceId {serviceId} ; :hasMaintenanceRemarks {remarks} .	SELECT "id", "serviceId", "serviceDate", "remarks" FROM "maintenance"
M3	:contact-{id} :hasCompanyName {company} ; :hasPhone {phone} ; :hasWebsite {website} .	SELECT "id", "company", "phone", "website" FROM "contact"
M4	:service/{Id} :hasServiceLabel {label} .	SELECT "id", "label" FROM "services"

Figure 6.15: Information patterns and mapping axioms for maintenance database

Data Extraction

An opportunity for the construction sector is to interrogate offshore steel inventories to ascertain suitable sections exist (DNS, AMEC and ZWS, 2016). For example, a sustainable design practitioner interested in suitable old gratings as drainage covers may interrogate the integrated data to find out the type and size of gratings available; any maintenance activities carried out on them; and the dates of the maintenance activities. Also, 3D scan positions close to the shortlisted grating elements can be connected for visual inspection. Figure 6.16 shows similar sample queries on the integrated data. Query Q7 aims to return all the relevant information about steel gratings on the boat deck of the hypothetical facility. The quest is to find out any maintenance activities carried out on them, and the dates the activities were performed. Also sought are the scan positions that are closest to them so that they can be visually assessed. Q7 answers this question by declaring that the grating on the boat deck (BD) is what is required from source S1. Grating objects are represented by their reference IDs. Using the common location BD, the RDF triples in source S3, relating to the required images, are matched. The closing part of the query extracts from source S2 the maintenance remarks and service date information for each steel item selected. It does this by matching the corresponding RDF triple patterns in S2 through the reference IDs of grating objects and maintenance occurrence. The result R7 of this query is presented in Figure 6.16.

Query Q8 in Figure 6.16 aims to list the types of physical objects in source S1 and describe them. This is where the generated reference data becomes useful. Q8 answers the question by using an RDF triple pattern from the reference data that declares that the items sought are a type of 'ClassOfInanimatePhysicalObject', an ISO 15926 Part 2 type entity. This pattern is then matched with a triple in S1 that connects the description of the items. The result of this query is presented as R8 in Figure 6.16. The last query, Q9, demonstrates that it is possible to return answers when parts of the data are missing. Q9 requests cut length and area values for steel items on the main deck of the hypothetical facility. The query extracts the result R9 by matching an RDF triple pattern, in source S1, that declares that the requested items are on the main deck (MD). This triple is then matched to RDF triple patterns in templates extracted from ISO 15926 – the project reference data. The area and cut length values are then optionally matched (SPARQL query function) through the blank nodes in the template patterns. Blank nodes are anonymous resources used for creating links between declared data objects (Cyganiak, Wood and Lanthaler, 2014).

Q7

```
#Query to extract maintenance and image data on boat deck gratings
PREFIX : <http://steel.waste.com/maintenance/>
PREFIX basens: <http://steel.waste.com/project/>
PREFIX qn1: <http://steel.waste.com/rdl/>
PREFIX qn3: <http://steel.waste.com/properties/>
PREFIX pcardl: <http://posccaesar.org/rdl/>

SELECT ?steel ?image ?remarks ?date
WHERE {
?steel qn3:hasLocation qn1:BD .
?steel pcardl:hasDesignation "Grating" .
basens:X-BD qn3:hasLocation qn1:BD .
basens:X-BD qn3:hasImage ?image .
SERVICE <http://localhost:8080/openrdf-sesame/repositories/Steel_Waste> {?steel :hasMaintenanceId ?id. ?id :hasMaintenanceRemarks ?remarks; :hasServiceDate ?date .}
}
```

R7

Steel (from S1)	Image (from S3)	Remarks (from S2)	Date (from S2)
S.03	"C:\Navisworks Data\Delta PP\Revised Truview with Model Superimposed\DATA\BOAT & SUBCELLAR DECK.htm"	"Replaced"	2011-07-05T00:00:00.0
S.03	"C:\Navisworks Data\Delta PP\Revised Truview with Model Superimposed\DATA\BOAT & SUBCELLAR DECK.htm"	"Corroded and needs replacement"	2017-09-18T00:00:00.0

Q8

```
# Query extracts the id of physical objects in data and their descriptions
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX pcardl: <http://posccaesar.org/rdl/>
PREFIX dm: <http://rds.posccaesar.org/2008/02/OWL/ISO-15926-2_2003#>

SELECT ?item ?description
WHERE {?item rdf:type dm:ClassOfInanimatePhysicalObject.
?item pcardl:hasDesignation ?description
}
```

R8

Item	Description
CHS168.3-5	"Circular Hollow Section"
CHS219.1-5.0	"Circular Hollow Section"
CHS323.9-5.6	"Circular Hollow Section"
CHS355.69-8.0	"Circular Hollow Section"
UB762-267-197	"Universal Beam"
PFC260-75-28	"Parrallel Flange Channel"
38-5SG	"Grating"

Q9

```
# Query extracts the geometric propertes of items on the main deck of the hypothetical facility
PREFIX qn4: <http://tpl.rdlfacade.org/data#>
PREFIX qn1: <http://steel.waste.com/rdl/>
PREFIX qn3: <http://steel.waste.com/properties/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dm: <http://rds.posccaesar.org/2008/02/OWL/ISO-15926-2_2003#>

SELECT ?itemNo ?cutLength ?area
WHERE {?itemNo qn3:hasLocation qn1:MD.
OPTIONAL {?z qn4:RBC4BE4B0BBAF47E4A0BEFFAE45B5C24B ?itemNo;
qn4:R0854474F25FA4F9699F0DA191B971B58 ?cutLength.}
OPTIONAL {?y qn4:R371BC00E162544848115CB372BB3DEF3 ?itemNo;
qn4:RD097D4A601E34341851FD3F991E93569 ?area.}
}
```

R9

ItemNo	CutLength	Area
M.03		"70 "
M.02	"15"	
M.01	"15"	
X-MD		

Figure 6.16: Sample queries and results for steel material assessment

6.2.4 Use Case Evaluation

This method is compared to the state-of-the-art tool for decommissioning waste assessment based on functions that are essential for developing the required inventory and extracting the correct information from it efficiently. Table 6.1 summarises the comparison of the two methods. The first issue is the development of an inventory that will contain all the required data for the process. This is going to be an arduous process if each line item in the state-of-the-art spreadsheet tool has to be manually populated. The proposed solution simplifies this process by preparing the sources and ingesting the data into a repository using data mapping techniques. This is particularly useful where large volumes of data are involved. It will eliminate any error that may be introduced in a manual process. Also, the mappings can be reused for additional data on the same project or for other projects that are similar.

Another advantage of the solution is that query results can be automatically ordered to suit a viewing preference. This is powered by SPARQL used for querying. On the other hand, the state-of-the-art solution has to be manually ordered. Some columns have to be hidden and may be forgotten for other cases if user does not pay careful attention. Another issue is that the filtering function in spreadsheets can eliminate rows of data if empty cells exist in the filtering column. Again, if the user does not pay careful attention, that information may be overlooked. In the proposed solution, SPARQL has the capability to query across empty value parts of data and return correct answers. An example is demonstrated with query Q9 in Figure 6.16.

Other benefits of the proposed solution relate to reasoning. The use of the state-of-the-art tool involves manually resolving meaning between terms to determine if they refer to same items e.g. ‘CHS 168.3X5.6’ and ‘CHS 219X5.0’ in Figure 6.11 are circular hollow sections. Also, the relationships between items have to be manually established and taken into consideration during the assessment process. The solution on the other hand uses an ontology of all the sources referenced to ISO 15926 for handling meaning and relationships between terms. For example, query Q8 in Figure 6.16 selects all types of physical objects in the case study data. To carry out the same function with the state-of-the-art tool, the user would have to go through each line item. The proposed solution also allows the specifying of exact details that are of interest. For example, query Q9 seeks only the information about items on the main deck of the hypothetical facility.

Table 6.1: Comparison of state-of-the-art tool and proposed solution

S/N	Issues	State of the Art Tool	Proposed Solution
1	Inventory Development	Manual inventory development process that is repetitive and therefore error prone.	Mapping files are used to migrate data from the sources. Method guarantees data in the sources will be ingested as-is. Also mappings created are reusable on similar tasks irrespective of the amount of data in the sources.
2	Concordanced Information	Method does not support automatic concordance table generation.	Supports automatic concordance table generation. The results are presented in the order desired.
3	Missing values	The filtering function in spreadsheets can eliminate rows of data relevant to a result where empty cells exist in the filtering column.	Makes use of SPARQL, a declarative query language that allows users to give broad instructions on data to retrieve. The engine underneath handles the retrieval process and returns the most complete result including where missing values exist.
4	Semantic Heterogeneity	Data elements with same meaning but represented differently have to be manually resolved in the spreadsheet.	Data elements that mean the same thing but represented differently are automatically resolved using ontology developed for the sources.
5	Level of Detail	This solution cannot reason about data with hierarchical relationships. The relationships have to be manually resolved.	The ontology developed in the solution handles the relationships between data entities and SPARQL can be used to extract information at desired level of detail.

6.2.5 Proposed Solution Limitations

One limitation of the proposed solution is that there is no opportunity to correct any wrong data already existing in the sources. This will be mapped as-is and may introduce error in the integrated data. Given this, integrity of the sources is critical. Another challenge is that some programming effort is required by the users to generate data mapping files and query the integrated data. This may be mitigated by leveraging open source tools that aid mapping and query generation e.g. Karma (Knoblock *et al.*, 2012). Lastly, enterprise grade severs are better suited for this type of technology deployment (Bilal, Oyedele, Akinade, *et al.*, 2016). This may not be suitable for small businesses as they may have budget constraints (Walsh, 2016). However, the cost should be offset against the current human approach that requires vast amount of costly error-prone human effort.

6.3 Summary

This chapter evaluated the proposed data integration framework for two use cases that are important to the stakeholders in the offshore decommissioning industry. The two examples showed that data in varying forms, and originating from multiple sources, can be easily integrated; and the integrity of the output data will remain consistent with the original sources. It also showed that the mappings, rules, queries and designed ontology can be revised and reused multiple times. The piping and equipment audit use case showed clearly that the precision and recall ability of the proposed solution is much better than the current manual approach being used. In the steel reuse use case, the proposed solution ensured semantic homogeneity while allowing the continued use of existing applications. This will ensure a low risk implementation and the little disruption to existing practice. The solution can also facilitate a more thorough assessment of offshore decommissioning wastes. Overall, the evaluation of the proposed data integration framework indicates potential efficiency and cost savings if adopted.

Chapter 7

Conclusion and Recommendations

This chapter summarises the result of this research. In addition, it highlights the benefits and limitations of the research. To conclude, it makes recommendations for further development of the work.

7.1 Summary of Thesis

The following provide reflection on the achievement of the research objectives and overall aim:

i. analysis of the offshore decommissioning process

This objective is targeted at the development of a business case for data integration in offshore decommissioning. Literature and unstructured interviews were used in understanding the traditional offshore decommissioning process, the resources available for offshore decommissioning waste management and the factors affecting the UK North Sea decommissioning industry. Thereafter, the result of a gap analysis conducted was summarised. The analysis showed that data integration technology applications in the process industry have focused on meeting the needs of other life cycle activities other than waste management, and the current best practice for decommissioning waste management still requires vast amount of human effort. This result fulfils the requirement of the first research objective.

ii. analysis of existing data integration approaches

This objective focused on the analysis of the requirements for the desired solution. After the set of requirements elicited from the stakeholders have been correctly validated and verified, solution approaches that address the identified needs were researched and reviewed. This process helped to identify the most appropriate approaches to address the needs. The most fitting options of the identified approaches were then matched to the stakeholders' requirements considering the constraints on the process and current practices for process plant data integration. The resulting analysis recommended an ontology-based technique for process industry data management. This result satisfies the second research objective.

iii. design of a data integration system with querying capability

For this objective, a thorough study of the technical resources (technologies and processes) required to construct solution artifacts was carried out. Thereafter, the resources were configured and tested in a prototype of the solution design using a simple case. Evaluation of the prototype's performance indicated that the application of open source tools based on Semantic Web technologies for data integration and querying is more efficient compared with the traditional manual data management approach. This outcome justifies the value of the design, satisfying the research objective.

iv. development of a life cycle data integration framework

This objective considered how best to handle lifecycle data of process plant items. From the unstructured interviews (see Table A.2), industry stakeholders underscored the application of ISO 15926 in any solution because it is the industry standard for lifecycle data integration. Consequently, consideration was given to identifying technologies that help in using the standard and the appropriate options were combined with the designed data integration system. The data integration framework that resulted was tested with a use case (Section 6.2) and the results were positive in comparison to a manual approach. This fulfills this research objective.

v. validation of developed data integration framework with prototypes

This objective focused on the evaluation of the prototype design using multiple use cases. The solution evaluation tasks were designed to assess, through measurement and analysis, the performance of and value delivered by the solution when applied in use cases of interest and compared to corresponding existing practices. They were also used to identify limitations of the proposed solution and recommend approaches for improvement. Application of the developed solution for offshore decommissioning waste management was demonstrated with two use cases involving piping, equipment and steel sections of a process plant facility. Results showed that the developed solution saved significant time and effort when compared to currently recommended manual approach. For example, two similar tasks carried out using the industry recommended tool took about 90 mins for an engineer to get results while the developed solution expended less than half of that time. Also, the implemented prototypes show that decommissioned items can be easily assessed for reuse. These results meet the requirements of this research objective.

Chapter 7: Conclusion and Recommendations

The overall aim of this research is to investigate the current decision-making environment in offshore decommissioning waste management and find ways of improving existing practices with the application of information technology. This study investigated the subject and identified that the management of waste data in the UK offshore decommissioning sector still involves a manual process. Consequently, it demonstrated how the use of Semantic Web technologies can help improve the process, thus fulfilling the research aim.

7.2 Solution Benefits

This study validates the assertion in Kang et al. (2013) and Schwalbe (2011) suggesting that business processes and methods achieve higher efficiency when redesigned with the use of information technology. Specific benefits of the proposed solution in this research include:

a) Use of industry standard: since the publication of the ISO 15926, very few studies exist on demonstrating use cases and implementation plans that can be easily adapted. Consequently, the developed solution demonstrated a simple way to use the standard for integrating heterogeneous process plant data from multiple sources. Using the standard will ensure shared understanding of a given data by all users.
b) Data quality check: the use of SPIN rules in the solution allows for automatic detection of data inconsistencies - something that is not possible in the manual inspection approach. By adding rules, inconsistencies are identified and eliminated at an early stage in the data processing.
c) Smooth data migration: use of mappings for data migration significantly reduces the effort required for extracting data from different sources in manual approaches. It also eliminates costly human errors that are prone to occur in manual approaches.
d) Task time reduction: while there is an initial setup cost with the developed approach, the tested prototypes indicated the cost is quickly recovered within just

a few tasks when compared to the current manual approaches. There is also a significant reduction in task duration as the number of repetition increases.

e) Resource reuse and extendibility: an advantage of this framework is that users having similar work processes will not need to start the creation of solution artifacts from scratch. They may use existing resources as-is, or adjust an existing set of templates (mappings, rules, queries and designed ontology). Either of the latter scenarios will result in significant reduction in the setup cost for a new project. If the same set of templates is used across multiple projects, the saving on the setup cost will further multiply. And when a set of templates is well tested, it will become an organisational process asset.
f) Connection to legacy database: The data integration framework contributed in this research will allow users to connect to relational database management systems, which is the traditional data repository in the oil and gas industry, thus following the path of least disruption to existing practice.

7.3 Potential Impacts of Work

This study has the potential to impact the following decommissioning process elements:

a) Supply chain: currently, screening out decommissioned items that are useful is a difficult task because information from multiple sources is required for such assessment. As a result, it is difficult to engage the supply chain and potential market outlets about the supply. Application of the solution proposed in this study is valuable because it can be applied for integrating heterogenous data for web platforms that enable searching. This will enable the searching of certified decommissioned items and reduce their time to market.
b) Job creation: a large percentage of decommissioned facilities is steel. Considering the quantity of steel materials involved in the decommissioning of process plants, there is a huge opportunity to extend their value based on circular economy principles. However, because of the limited steel recycling capacity in the UK and the preference of asset owner for recycling, most of the materials are sent overseas for processing. The owners' preference for recycling is, in part, due to the challenges associated with assessing life cycle information of steel materials for reuse. Application of the proposed solution in this study will enable assessments

for reuse. Such assessments have the potential of changing the course of decommissioned steel by finding applications that will use them in the UK.

c) Decommissioning cost: the solution developed in this research can help to improve time and cost efficiencies of the offshore decommissioning process. It will improve time efficiency by reducing the effort spent in making sense of heterogenous data. It will also reduce costly human errors and save on paid man-hours. These, in turn, will result in compressed work schedules, allowing for quicker turnaround.

d) Small businesses: this study demonstrated the use of open source tools for data integration. It also provided a data integration architecture with multiple use case templates that can be adopted. These will be useful resources for small businesses with limited budget for proprietary applications (Dainty *et al.*, 2017), but in need of data integration tools.

e) Environment: decommissioning in the UK must comply with relevant environmental regulations (UK Gov, 2008). As a result, decommissioned materials must be managed appropriately in an environmentally responsible way. The three methods stipulated for managing decommissioning waste include disposal on land, recycling and reuse (DECC, 2011). Disposal on land (degradability issues) and recycling (carbon footprint) have higher environmental implications than reuse, so the latter is best for the environment. However, determining whether decommissioned material can be reused requires reliability assessment which depends on the associated life cycle data of the items. Consequently, the solution from this study is relevant because it can be applied in integrating life cycle data of the items and answering questions about them. This will help to identify materials that can be put back to use and by implication, benefit the environment.

7.4 Research Limitations

The main limitations to this research are related to the following:

a) Acceptance and evaluation Criteria for the developed solution: achieving agreement on acceptance and evaluation criteria for different needs among diverse stakeholders was challenging because some organisations have ample resources at their disposal and were accustomed to investing in proprietary tools configured to their needs. So, while businesses with limited resources articulated their challenges very well, those with ample resources seemed comfortable with changing their work processes based on new tools. Consequently, solutions developed seem to address the problem from the view of small businesses.

b) Functional decomposition of offshore decommissioning process: the offshore decommissioning process is a complex process and cannot be completely represented by the high-level model used in this research. This is because some interactions between its elements can cause emergent options with special considerations. Exploring all the possible scenarios can be a challenging and time-consuming task. As a result, this research is aligned to the most traditional offshore decommissioning method which may not capture important considerations in alternative methods.

c) Technology for developed solution: one challenge of the proposed solution is that some programming effort is required by the users to generate data mapping files and query the integrated data. This is in part due to the limited focus on human factors. Nonetheless, there is potential for mitigating this by leveraging open source tools that aid mapping and query generation e.g. Karma (Knoblock *et al.*, 2012) and VIOLA (Cao, Lin and Zhu, 2018) . Another technology challenge is that enterprise grade severs are better suited for this type of technology deployment (Bilal, Oyedele, Akinade, *et al.*, 2016) which may not be suitable for small businesses because of budget constraints (Walsh, 2016). However, the cost should be offset against the current human approach, that requires vast amount of costly error-prone human effort, in deciding. Opportunities with pay-per-use cloud computing (Cao, Lin and Zhu, 2018) may also be valuable in mitigating the cost.

7.5 Recommendations for Future Work

While the result of the proposed framework indicates potential productivity gains for data projects, its declared benefits need to be further tested by more substantive use case applications. The framework should be applied by engineers working on similar tasks in the process industry and evaluations on productivity, scalability and cost benefits carried out.

Future research effort should also seek to further reduce the setup cost of the proposed framework to lower the barrier for its application. This could mean developing solutions that integrate the mapping process, data storage and querying interface in one place. Such solutions should also automate all the steps in the process for ease of use by non-programmers.

For decommissioning waste management, the next step should seek to build a graphical user interface that will exploit the querying capabilities of the developed framework to support reliability assessment of specific decommissioned items. Such solutions will be valuable for decommissioning supply chain, enabling the searching of certified decommissioned items that can be reused, thus reducing the time of such items to market. It will be valuable to see an online application in this domain that is able to synthesise available life cycle data about a decommissioned item, carry out prescriptive analysis and recommend future uses to which it can be applied. In addition, such application should allow users to create accounts and add associated life cycle data of items for appraisal, display and sales.

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*. Pergamon, 40(14), pp. 5563–5577. doi: 10.1016/j.eswa.2013.04.027.

Aguilar-Savén, R. S. (2004) ‘Business process modelling: Review and framework’, *International Journal of Production Economics*, 90(2), pp. 129–149. doi: 10.1016/S0925-5273(03)00102-6.

Akinyemi, A. G. (2018a) *Data Mapping Using dot15926 Editor*, *Zenodo*. doi: 10.5281/zenodo.1405561.

Akinyemi, A. G. (2018b) *Ontology-Based Data Access Using Ontop*, *Zenodo*. doi: 10.5281/zenodo.1439101.

Akinyemi, A. G. (2018c) *Process plant data integration and querying*, *Zenodo*. doi: 10.5281/zenodo.1321818.

Al-Mamary, Y. H., Shamsuddin, A. and Aziati, N. (2014) ‘The Role of Different Types of Information Systems In Business Organizations : A Review’, *International Journal of Research*, 1(7), pp. 1279–1286.

Almendros-Jiménez, J. M., Becerra-Terón, A. and Cuzzocrea, A. (2017) ‘Syntactic and semantic validation of SPARQL queries’, in *SAC ’17 Proceedings of the Symposium on Applied Computing*. Morocco, pp. 349–352.

Ariosh (2018) *Ariosh*. Available at: http://www.ariosh.com (Accessed: 3 May 2018).

ARUP, Scottish Enterprise and DNS (2014) *Decommissioning in the North Sea*. Available at: https://www.raeng.org.uk/publications/reports/decommissioning-in-the-north-sea (Accessed: 16 August 2017).

Banias, G., Achillas, C., Vlachokostas, C., Moussiopoulos, N. and Papaioannou, I. (2011) ‘A web-based Decision Support System for the optimal management of construction and demolition waste’, *Waste Management*. Elsevier Ltd, 31(12), pp. 2497–2502. doi: 10.1016/j.wasman.2011.07.018.

Bayer, B. and Marquardt, W. (2004) ‘Towards integrated information models for data’, *Computers and Chemical Engineering*, 28, pp. 1249–1266. doi: https://doi.org/10.1016/j.compchemeng.2003.08.008.

Bermejo, J. (2007) *A simplified guide to create an ontology*. Available at: http://tierra.aslab.upm.es/documents/controlled/ASLAB-R-2007-004.pdf (Accessed: 6 February 2018).

Berners-Lee, T., Hendler, J. and Lassila, O. (2001) ‘The Semantic Web’, *Scientific American*, 284(5), pp. 34–43.

Bilal, M., Oyedele, L. O., Akinade, O. O., Ajayi, S. O., Alaka, H. A., Owolabi, H. A., Qadir, J., Pasha, M. and Bello, S. A. (2016) ‘Big data architecture for construction waste analytics (CWA): A conceptual framework’, *Journal of Building Engineering*, 6, pp. 144–156. doi: 10.1016/j.jobe.2016.03.002.

Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., Owolabi, H. A., Alaka, H. A. and Pasha, M. (2016) 'Big Data in the construction industry: A review of present status, opportunities, and future trends', *Advanced Engineering Informatics*. Elsevier Ltd, 30(3), pp. 500–521. doi: 10.1016/j.aei.2016.07.001.

Bioregional (2008) *Reclaimed building products guide*. Available at: http://www.wrap.org.uk/sites/files/wrap/Reclaimed building products guide.pdf (Accessed: 16 August 2017).

Bittner, K. (2002) *Use Case Modeling*. Boston, MA: Addison-Wesley Longman Publishing Co.

Borgo, S. (2014) 'An ontological approach for reliable data integration in the industrial domain', *Computers in Industry*. Elsevier B.V., 65(9), pp. 1242–1252. doi: 10.1016/j.compind.2013.12.010.

Borgo, S. and Masolo, C. (2009) 'Ontologies', in Staab, S. and Studer, R. (eds) *Handbook on Ontologies*. Berlin, Germany: Springer, pp. 361–381. doi: 10.1007/978-3-540-92673-3.

Bravo, C., Castro, J. A., Saputelli, L., Ríos, A., Aguilar-Martin, J. and Rivas, F. (2011) 'An implementation of a distributed artificial intelligence architecture to the integrated production management', *Journal of Natural Gas Science and Engineering*, 3(6), pp. 735–747. doi: 10.1016/j.jngse.2011.08.002.

Brickley, D. and Guha, R. V. (2014) *RDF Schema 1.1*, *W3C Recommendation*. Available at: https://www.w3.org/TR/rdf-schema/ (Accessed: 6 February 2018).

Bruijn, J. de and Welty, C. (2013) *RIF RDF and OWL Compatibility (Second Edition)*, *W3C Recommendation*. Available at: https://www.w3.org/TR/2013/REC-rif-rdf-owl-20130205/ (Accessed: 6 February 2018).

Bryman, A. and Bell, E. (2015) *Business Research Methods*. 4th edn. New York, NY: Oxford University Press.

Bureau Veritas (2011) *Decommissioning on the UK Continental Shelf - an overview of regulations*. Available at: http://www.bureauveritas.co.uk/QALDNKHI/BV_Decommissioning+document_low+res-v3.pdf (Accessed: 16 August 2017).

Cabantous, L. and Gond, J.-P. (2010) 'Rational Decision Making as Performative Praxis: Explaining Rationality's Éternel Retour', *Organisation Science*, 22(3), pp. 573–586.

Calvanese, D., Cogrel, B., Komla-Ebri, S., Kontchakov, R., Lanti, D., Rezk, M., Rodriguez-Muro, M. and Xiao, G. (2017) 'Ontop: Answering SPARQL queries over relational databases', in *Semantic Web*, pp. 471–487. doi: 10.3233/SW-160217.

Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A., Rodriguez-Muro, M. and Rosati, R. (2009) 'Ontologies and databases: The dl-lite approach', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5689 LNCS, pp. 255–356. doi: 10.1007/978-3-642-03754-2_7.

Cao, N., Lin, C. and Zhu, Q. (2018) 'Voila: Visual Anomaly Detection and Monitoring with

Streaming Spatiotemporal Data', *IEEE Transactions on Visualization and Computer Graphics*, 24(1), pp. 23–33.

Chamanara, J., König-Ries, B. and Jagadish, H. V (2017) 'QUIS: In-Situ Heterogeneous Data Source Querying', *Proceedings of the VLDB Endowment*, 10(12), pp. 1877–1880. doi: 10.14778/3137765.3137798.

Chen, W., Li, J., Ma, J., Conradi, R., Ji, J. and Liu, C. (2008) 'An empirical study on software development with open source components in the chinese software industry', in *Software Process: Improvement and Practice*, pp. 89–100. doi: 10.1002/spip.361.

Christou, M. and Konstantinidou, M. (2012) *Safety of offshore oil and gas operations: Lessons from past accident analysis*. Available at: http://publications.jrc.ec.europa.eu/repository/bitstream/JRC77767/offshore-accident-analysis-draft-final-report-dec-2012-rev6-online.pdf (Accessed: 16 August 2017).

Clark, J. W. (2010) *Business Intelligence and Decision Making: Understanding BI as a Theory-performing Discipline of Decision Improvement*, *Proceedings of JAIS Theory Development Workshop. Working Papers on Information Systems*.

Coleman, D. (1998) *A Use Case Template: draft for discussion*. Available at: http://www.engr.sjsu.edu/~fayad/current.courses/cmpe202-Fall2009/docs/lecture3/CmpE202-22-UC-Template-Ex-L3-3g.pdf.

Connolly, T. M. and Begg, C. E. (2005) *Database systems : a practical approach to design, implementation, and management*. 4th edn. Addison-Wesley.

Covaliu, Z. (2001) *Decision Analysis: Concepts, Tools and Promise A Fair Isaac White Paper*. Available at: http://www.edmblog.com/weblog/files/decision_analytics_white_paper.pdf (Accessed: 12 February 2018).

CrowdFlower (2016) *CrowdFlower Data Science Report*. Available at: http://visit.crowdflower.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf (Accessed: 6 February 2018).

Curé, O. and Blin, G. (2015) *RDF database systems : triples storage and SPARQL query processing*. Elsevier.

Cyganiak, R., Wood, D. and Lanthaler, M. (2014) *RDF 1.1 Concepts and Abstract Syntax*, *W3C Recommendation*. Available at: https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/ (Accessed: 6 February 2018).

D3 Consulting Limited (2017) *Waste Management in Offshore Decommissioning*, *Late Life Planning Portal*. Available at: https://l2p2.net/SitePages/Tool-Download.aspx?DocId=445&Version=4.0 (Accessed: 18 October 2017).

Dabbish, L., Stuart, C., Tsay, J. and Herbsleb, J. (2012) 'Social coding in GitHub: transparency and collaboration in an open software repository', in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, pp. 1277–1286. doi: 10.1145/2145204.2145396.

Dainty, A., Leiringer, R., Fernie, S. and Harty, C. (2017) 'BIM and the small construction firm: a critical perspective', *Building Research and Information*. Taylor & Francis, 45(6), pp. 696–709. doi: 10.1080/09613218.2017.1293940.

Dalgic, Y., Lazakis, I. and Turan, O. (2013) 'Vessel charter rate estimation for offshore wind O&M activities', in Soares, C. G. and Peña, F. L. (eds) *Developments in Maritime Transportation and Exploitation of Sea Resources*. CRC Press, pp. 899–907.

Das, S., Sundara, S. and Cyganiak, R. (2012) *R2RML: RDB to RDF Mapping Language*, *W3C Recommendation*. Available at: https://www.w3.org/TR/r2rml/ (Accessed: 6 February 2018).

Davis, J. and Goadrich, M. (2006) 'The relationship between Precision-Recall and ROC curves', *Proceedings of the 23rd international conference on Machine learning - ICML '06*, pp. 233–240. doi: 10.1145/1143844.1143874.

DECC (2011) *Guidance Notes: Decommissioning of Offshore Oil and Gas Installations and Pipelines under the Petroleum Act 1998*. Available at: https://pdfs.semanticscholar.org/67da/9f8a672cc7768748569e37be2c551dce5b08.pdf (Accessed: 16 August 2017).

Deloitte (2015) *Insights: The impact of plummeting crude oil prices on company finances*. Available at: http://www2.deloitte.com/ng/en/pages/energy-and-resources/articles/crude-awakening-the-impact-of-plummeting-crude-oil-prices-on-company-finances.html# (Accessed: 16 August 2017).

Denno, P. and Palmer, M. (2013) 'Modeling and Conformance Testing for the Engineering Information Integration Standard ISO 15926', pp. 1–22. doi: dx.doi.org/10.6028/NIST.IR.7917.

Despa, M. L. (2014) 'Comparative Study on Software Development Methodologies', *Database Systems Journal*, 5(3), pp. 37–56. doi: 10.1109/MAHC.1983.10102.

Devold, H. (2013) *Oil and gas production handbook An introduction to oil and gas production, transport, refining and petrochemical industry*. Available at: https://library.e.abb.com/public/34d5b70e18f7d6c8c1257be500438ac3/Oil and gas production handbook ed3x0_web.pdf (Accessed: 16 August 2017).

Dimou, A., Sande, M. Vander, Colpaert, P., Verborgh, R., Mannens, E. and Van De Walle, R. (2014) 'RML: A generic language for integrated RDF mappings of heterogeneous data', in *Proc. of the 7th LDOW workshop*.

Djamaluddin, B., International, P., Mohammed, S. and Fahd, K. (2018) 'Survey of Big Data New SQL Database Systems for Real-Time Data Transfer in the Upstream Oil and Gas Industry', in *SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition*. Dammam. doi: 10.2118/192412-MS.

DNS (2017) *Late Life Planning Portal*. Available at: https://l2p2.net/SitePages/Home.aspx (Accessed: 16 August 2017).

DNS, AMEC and ZWS (2016) *Decommissioned Steel Reuse in Construction*. Aberdeen.

DNS, ZWS and ABB (2015) *Offshore Oil and Gas Decommissioning*, *Offshore Oil and Gas Decommissioning*. Available at:

https://library.e.abb.com/public/d689c2f70f0c447586610ac566c9aa7e/ABB-Offshore-Oil-and-Gas-Decommissioning-2015.pdf (Accessed: 16 August 2017).

Doan, A., Halevy, A. and Ives, Z. (2012a) 'Introduction', in *Principles of Data Integration*. Elsevier, pp. 1–18. doi: 10.1016/B978-0-12-416044-6.00001-6.

Doan, A., Halevy, A. and Ives, Z. (2012b) 'Ontologies and Knowledge Representation', in *Principles of Data Integration*, pp. 325–344. doi: 10.1016/B978-0-12-416044-6.00012-0.

Doan, A., Halevy, A. and Ives, Z. (2012c) *Principles of Data Integration*. Elsevier.

Ebrahimipour, V. and Yacout, S. (2015) 'Ontology-Based Knowledge Platform to Support Equipment Health in Plant Operations', in Ebrahimipour, V. and Yacout, S. (eds) *Ontology Modeling in Physical Asset Integrity Management*. Springer, Cham, pp. 221–255. doi: 10.1007/978-3-319-15326-1_8.

Ekins, P., Vanner, R. and Firebrace, J. (2005) *Decommissionig of offshore oil and gas facilities: Decomissioning Scenarios: A Comparative Assessment Using Flow Analysis*. Available at: http://www.psi.org.uk/docs/2005/UKOOA/Decommissioning-Working paper.pdf (Accessed: 16 August 2017).

Ellen MacArthur Foundation (2015) *Towards a Circular Economy: Business Rationale for an Accelerated Transition*. Available at: https://www.ellenmacarthurfoundation.org/assets/downloads/TCE_Ellen-MacArthur-Foundation_9-Dec-2015.pdf (Accessed: 6 February 2018).

Embury, S. M., Brandt, S. M., Robinson, J. S., Sutherland, I., Bisby, F. A. and Gray, W. A. (2001) 'Adapting integrity enforcement techniques for data reconciliation', *Information Systems*, 26(8), pp. 657–689. doi: https://doi.org/10.1016/S0306-4379(01)00044-8.

Erwin, K. (2011) 'Part G: Probability, Statistics', in 10th (ed.) *Advanced Engineering Mathematics*. Wiley, pp. 1013–1014.

Executive and Health Saftey (2014) *Offshore Oil & Gas Sector Strategy 2014 To 2017*. Available at: http://www.hse.gov.uk/offshore/offshore-strategic-context.pdf (Accessed: 16 August 2017).

Fawcett, T. (2006) 'An introduction to ROC analysis', *Pattern Recognition Letters*, 27(8), pp. 861–874. doi: 10.1016/j.patrec.2005.10.010.

Fiatech (2011) *An Introduction to ISO 15926*. Available at: http://www.fiatech.org/images/stories/techprojects/project_deliverables/iso-intro-ver1.pdf (Accessed: 6 February 2018).

Fiatech, PCA and USPI (2015) *Building the Semantic Web for the Process Industries using RDF, OWL and SPARQL for the Integration, Sharing, Exchange, and Hand-over of distributed Plant Lifecycle Information on the basis of ISO 15926*. Available at: http://www.infowebml.ws/index.htm (Accessed: 6 February 2018).

Fiorentini, X., Paviot, T., Fortineau, V., Goblet, J.-L. and Lamouri, S. (2013) 'Modeling nuclear power plants engineering data using ISO 15926', in *Proceedings of 2013 International Conference on Industrial Engineering and Systems Management (IESM)*. Rabat: IEEE, pp. 1–6.

Gallaher, M. P., O'Connor, A. C., Jr., D. and J. L., & Gilday, L. T. (2004) *Cost analysis of inadequate interoperability in the U.S. capital facilities industry*. Available at: http://nvlpubs.nist.gov/nistpubs/gcr/2004/NIST.GCR.04-867.pdf.

Gielingh, W. (2008) 'An assessment of the current state of product data technologies', *CAD Computer Aided Design*, 40(7), pp. 750–759. doi: 10.1016/j.cad.2008.06.003.

Golzarpoor, B., Haas, C. T., Rayside, D., Kang, S. and Weston, M. (2018) 'Improving construction industry process interoperability with Industry Foundation Processes (IFP)', *Advanced Engineering Informatics*. Elsevier, 38(September), pp. 555–568. doi: 10.1016/j.aei.2018.09.001.

Gray, A. J. G., Groth, P., Loizou, A., Askjaer, S., Brenninkmeijer, C., Burger, K., Chichester, C., Evelo, C. T., Goble, C., Harland, L., Pettifer, S., Thompson, M., Waagmeester, A. and Williams, A. J. (2014) 'Applying linked data approaches to pharmacology: Architectural decisions and implementation', *Semantic Web*, 5(2), pp. 101–113. doi: 10.3233/SW-2012-0088.

Gregor, S. (2006) 'The Nature of Theory in Information Systems', *Management Information Systems*, 30(3), pp. 611–642.

Groves, A. (2016) 'Beyond Excel: how to start cleaning data with OpenRefine', *Multimedia Information and Technology*, 42(2), pp. 18–22.

Hanseth, O. and Lyytinen, K. (2004) 'Theorizing about the Design of Information Infrastructures : Design Kernel Theories and Principles', *Proceedings of JAIS Theory Development Workshop. Working Papers on Information Systems*, 4(12), pp. 207–241.

Hauge, Ø., Ayala, C. and Conradi, R. (2010) 'Adoption of open source software in software-intensive organizations – A systematic literature review', *Information and Software Technology*. Elsevier, 52(11), pp. 1133–1154. doi: 10.1016/J.INFSOF.2010.05.008.

Hevner, A. R., March, S. T., Park, J. and Ram, S. (2004) 'Design science in information systems research', *MIS Quaterly*, 28(1), pp. 75–105.

HM Revenue & Customs (2012) *Annual Report and Accounts 2011-12*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/89198/annual-report-accounts-1112.pdf (Accessed: 16 August 2016).

Horakova, M., Skalska, H., Olszak, C. M., Ziemba, E., Fitriana, R., Djatna, T., Kursan, I., Mihić, M., Albescu, F., Pugna, I. and Paraschiv, D. (2013) 'Business Intelligence and Implementation in a Small Enterprise.', *Journal of Systems Integration*, 4(2), pp. 50–62. doi: 18042724.

Horridge, M., Knublauch, H., Rector, A., Stevens, R., Wroe, C., Jupp, S., Moulton, G., Drummond, N. and Brandt, S. (2011) *A Practical Guide To Building OWL Ontologies Using Protégè 4 and CO-ODE Tools Edition 1.3*. The University of Manchester. Available at: http://mowl-power.cs.man.ac.uk/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_3.pdf.

Horridge, M. and Patel-Schneider, P. F. (2012) *OWL 2 Web Ontology Language Manchester Syntax*, *W3C Working Group Note*. Available at:

https://www.w3.org/TR/owl2-manchester-syntax/ (Accessed: 6 February 2018).

Horrocks, I. and Patel-Schneider, P. (2011) ‘Knowledge Representation and Reasoning on the Semantic Web: OWL’, in Domingue, J., Fensel, D., and Hendler, J. A. (eds) *Handbook of Semantic Web Technologies*. Springer, pp. 365–398. doi: https://doi.org/10.1007/978-3-540-92913-0_9.

HSE Energy Division (2014) *Offshore oil and gas: Offshore Strategic context. Retrieved from Health and Safety Executive*. Available at: http://www.hse.gov.uk/offshore/offshore-oil-and-gas.pdf (Accessed: 16 August 2017).

IIBA (2015) *A Guide to the Business Analysis Body of Knowledge (BABOK)*. Toronto: International Institute of Business Analysis.

ISO (2003) *Industrial automation systems and integration — Integration of life-cycle data for process plants including oil and gas production facilities — Part 2: Data model*. First edit. Geneva. doi: 10.1109/IEEESTD.2007.4288250.

ISO (2004) *ISO 10303-11:2004 -- Industrial automation systems and integration -- Product data representation and exchange -- Part 11: Description methods: The EXPRESS language reference manual*. 2nd edn. Available at: https://www.iso.org/standard/38047.html (Accessed: 31 December 2018).

ISO (2007a) *ISO 10303-28:2007 -- Industrial automation systems and integration -- Product data representation and exchange -- Part 28: Implementation methods: XML representations of EXPRESS schemas and data, using XML schemas*. 1st edn. Available at: https://www.iso.org/standard/40646.html (Accessed: 31 December 2018).

ISO (2007b) *ISO 12006-3:2007 -- Building construction -- Organization of information about construction works -- Part 3: Framework for object-oriented information*. 1st edn. Available at: https://www.iso.org/standard/38706.html (Accessed: 31 December 2018).

ISO (2012) *ISO 29481-2:2012 -- Building information models -- Information delivery manual -- Part 2: Interaction framework*. 1st edn. Available at: https://www.iso.org/standard/55691.html (Accessed: 31 December 2018).

ISO (2016) *ISO 10303-21:2016--Industrial automation systems and integration -- Product data representation and exchange -- Part 21: Implementation methods: Clear text encoding of the exchange structure*. 3rd edn. Available at: https://www.iso.org/standard/63141.html (Accessed: 31 December 2018).

ISO (2018a) *ISO 10303-42:2018 -- Industrial automation systems and integration -- Product data representation and exchange -- Part 42: Integrated generic resource: Geometric and topological representation*. 5th edn. Available at: https://www.iso.org/standard/76128.html (Accessed: 31 December 2018).

ISO (2018b) *ISO 16739-1:2018 -- Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries -- Part 1: Data schema*. 1st edn. Available at: https://www.iso.org/standard/70303.html (Accessed: 31 December 2018).

Jamshed, S. (2014) ‘Qualitative research method-interviewing and observation’, *Journal of Basic and Clinical Pharmacy*. 3rd edn, 5(4), pp. 87–88. doi: 10.4103/0976-0105.141942.

Kang, Y., William, J. and Stephen, P. (2013) ‘Value of IT: Indirect Impact of IT on Construction Project Performance Via Best Practices’, *Automation in Construction*, 35, pp. 383–396. doi: https://doi.org/10.1016/j.autcon.2013.05.011.

Kaur, N. and Aggarwal, H. (2017) ‘Evaluation of Information Retrieval Based Ontology Development Editors for Semantic Web’, *International Journal of Modern Education and Computer Science*, 9(7), pp. 63–73.

Kim, B. C., Teijgeler, H., Mun, D. and Han, S. (2011) ‘Integration of distributed plant lifecycle data using ISO 15926 and Web services’, *Annals of Nuclear Energy*. Elsevier Ltd, 38(11), pp. 2309–2318. doi: 10.1016/j.anucene.2011.07.033.

Kitchin, R. (2014) ‘Big data’, in Rojek, R., DIckens, K., Haw, K., and Rose, J. (eds) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE, pp. 67–79.

Knap, T., Hanecák, P., Klímek, J., Mader, C. and Necaský, M. (2018) ‘UnifiedViews: An ETL Tool for RDF Data Management’, *Semantic Web*. doi: https://doi.org/10.3233/SW-180291.

Knoblock, C. A., Szekely, P., Ambite, J. L., Goel, A., Gupta, S., Lerman, K., Muslea, M., Taheriyan, M. and Mallick, P. (2012) ‘Semi-Automatically Mapping Structured Sources into the Semantic Web’, in *Proc. of the 9th Extended SemanticWeb Conf. (ESWC)*, pp. 375–390. doi: doi:10.1007/978-3-642-30284-8_32.

Knublauch, H. (2013) *SPIN - SPARQL Syntax*, *W3C Member Submission*. Available at: http://spinrdf.org/sp.html (Accessed: 25 September 2017).

Laud, A. (2013) ‘Interoperability between IFC’s (ISO 16739) and ISO 15926’, pp. 1–3. Available at: http://iringtoday.com/wordpress/wp-content/uploads/2012/01/IFCs-and-ISO-15926.pdf.

Lee, D.-Y., Chi, H., Wang, J., Wang, X. and Park, C.-S. (2016) ‘A linked data system framework for sharing construction defect information using ontologies and BIM environments’, *Automation in Construction*. Elsevier, 68, pp. 102–113. doi: 10.1016/J.AUTCON.2016.05.003.

Lee, S., Han, S. and Mun, D. (2012) ‘Integrated management of facility, process, and output: data model perspective’, *Science China Information Sciences*, 55(5), pp. 994–1007. doi: 10.1007/s11432-012-4555-1.

Li, Q. and Chen, Y. (2009) ‘Data Flow Diagram’, in *Modeling and Analysis of Enterprise and Information Systems*. Springer, Berlin, Heidelberg, pp. 85–97. Available at: https://doi.org/10.1007/978-3-540-89556-5_4.

Li, Y. and Zhang, X. (2013) ‘Web-based construction waste estimation system for building construction projects’, *Automation in Construction*. Elsevier B.V., 35, pp. 142–156. doi: 10.1016/j.autcon.2013.05.002.

Linetzki, A. (2012) ‘Analyzing decision tables’, *TestingExperience Magazine*, March, pp. 1–11. Available at: http://www.imbus.cn/upFile/uploadfiles/201295949378154.pdf.

Marsan, J., Paré, G. and Beaudry, A. (2012) ‘Adoption of open source software in organizations: A socio-cognitive perspective', *The Journal of Strategic Information Systems*. North-Holland, 21(4), pp. 257–273. doi: 10.1016/J.JSIS.2012.05.004.

McGrath, C. (2001) 'Waste minimisation in practice', *Resources, Conservation and Recycling*, 32(3–4), pp. 227–238. doi: 10.1016/S0921-3449(01)00063-5.

McMurry, J. A., Juty, N., Blomberg, N., Burdett, T., Conlin, T., Conte, N., Courtot, M., Deck, J., Dumontier, M., Fellows, D. K., Gonzalez-Beltran, A., Gormanns, P., Grethe, J., Hastings, J., Hériché, J. K., Hermjakob, H., Ison, J. C., Jimenez, R. C., Jupp, S., Kunze, J., Laibe, C., Le Novère, N., Malone, J., Martin, M. J., McEntyre, J. R., Morris, C., Muilu, J., Müller, W., Rocca-Serra, P., Sansone, S. A., Sariyar, M., Snoep, J. L., Soiland-Reyes, S., Stanford, N. J., Swainston, N., Washington, N., Williams, A. R., Wimalaratne, S. M., Winfree, L. M., Wolstencroft, K., Goble, C., Mungall, C. J., Haendel, M. A. and Parkinson, H. (2017) 'Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data', *PLoS Biology*, 15(6), pp. 1–27. doi: 10.1371/journal.pbio.2001414.

Miljkovic, D. (2005) 'Rational choice and irrational individuals or simply an irrational theory: A critical review of the hypothesis of perfect rationality', *The Journal of Socio-Economics*. North-Holland, 34(5), pp. 621–634. doi: 10.1016/J.SOCEC.2003.12.031.

Motik, B., Grau, B. C., Horrocks, I., Wu, Z., Fokoue, A. and Lutz, C. (2012) *OWL 2 Web Ontology Language*, *W3C Recommendation*. Available at: https://www.w3.org/TR/owl2-profiles/ (Accessed: 6 February 2018).

Musen, M. A. (2015) 'The Protégé project: A look back and a look forward', *Association of Computing Machinery Specific Interest Group in Artificial Intelligence*, 1(4). doi: 10.1145/2557001.25757003.

Mutis, I. and Issa, R. R. A. (2012) 'Framework for Semantic Reconciliation of Construction Project Information', *Journal of Information Technology in Construction (ITcon)*, 17(2), pp. 1–24.

Nasdaq (2017) *Crude Oil*. Available at: http://www.nasdaq.com/markets/crude-oil.aspx?timeframe=5y (Accessed: 16 August 2017).

Nextspace (2018) *XMpLant*. Available at: http://www.nextspace.co.nz/products-and-services/solutions/xmplant/ (Accessed: 31 December 2018).

Niknam, M. and Karshenas, S. (2015) 'Integrating distributed sources of information for construction cost estimating using Semantic Web and Semantic Web Service technologies', *Automation in Construction*. Elsevier, 57, pp. 222–238. doi: 10.1016/J.AUTCON.2015.04.003.

Novák, P., Serral, E., Mordinyi, R. and Šindelář, R. (2015) 'Integrating heterogeneous engineering knowledge and tools for efficient industrial simulation model support', *Advanced Engineering Informatics*, 29(3), pp. 575–590. doi: 10.1016/j.aei.2015.05.001.

Novák, P., Šindelář, R. and Mordinyi, R. (2014) 'Integration framework for simulations and SCADA systems', *Simulation Modelling Practice and Theory*, 47, pp. 121–140. doi: 10.1016/j.simpat.2014.05.010.

O'Connor, M. J., Halaschek-Wiener, C. and Musen, M. A. (2010) 'Mapping Master: A

Flexible Approach for Mapping Spreadsheets to OWL', in *9th International Semantic Web Conference*. Shanghai.

Oil and Gas UK (2014) *Economic Report 2014*. Available at: http://oilandgasuk.co.uk/2014-economic-report.cfm (Accessed: 16 August 2017).

Oil and Gas UK (2015a) *Activity Survey 2015*. Available at: http://oilandgasuk.co.uk/wp-content/uploads/2015/07/EC044.pdf (Accessed: 16 August 2017).

Oil and Gas UK (2015b) *Decommissioning Insight 2015*. Available at: http://oilandgasuk.co.uk/wp-content/uploads/2015/11/Decommissioning-Insight-2015-updated.pdf (Accessed: 16 August 2017).

Olubunmi, F. and Ward, R. (2017) *The End of Handover*. Fiatech. Available at: https://www.youtube.com/watch?v=wfVP7snMis8 (Accessed: 6 February 2018).

Omar, T. and Nehdi, M. L. (2016) 'Data acquisition technologies for construction progress tracking', *Automation in Construction*. Elsevier, 70, pp. 143–155. doi: 10.1016/J.AUTCON.2016.06.016.

Onwuegbuzie, A. J., Johnson, R. B. and Collins, K. M. T. (2009) 'Call for mixed analysis: A philosophical framework for combining qualitative and quantitative approaches', *International Journal of Multiple Research Approaches*, 3(2), pp. 114–139. doi: 10.5172/mra.3.2.114.

Pauwels, P., Zhang, S. and Lee, Y. C. (2017) 'Semantic web technologies in AEC industry: A literature overview', *Automation in Construction*. Elsevier B.V., 73, pp. 145–165. doi: 10.1016/j.autcon.2016.10.003.

PMI (2013) *A Guide to the Project Management Body of Knowledge*. Newtown Square: Project Management Institute. doi: 10.1002/pmj.21345.

Pollack, S. L. (1963) *Analysis of the decision rules in decision tables*. Available at: http://www.dtic.mil/dtic/tr/fulltext/u2/407719.pdf (Accessed: 3 June 2018).

Power, D. J., Sharda, R. and Burstein, F. (2015) 'Decision Support Systems', *Wiley Encyclopedia of Management*. doi: DOI: 10.1002/9781118785317.weom070211.

Protégé Project (2017) *Cellfie*. Available at: https://github.com/protegeproject/cellfie-plugin (Accessed: 20 August 2018).

Prud'hommeaux, E., Buil-Aranda, C., Seaborne, A., Polleres, A., Feigenbaum, L. and Williams, G. T. (2013) *SPARQL 1.1 Federated Query*, *W3C Recommendation*. Available at: https://www.w3.org/TR/sparql11-federated-query/ (Accessed: 26 October 2016).

Qady, M. Al and Kandil, A. (2012) 'Document Discourse for Managing Construction Project Documents', *Journal of Computing in Civil Engineering*, 27(5), pp. 466–475.

Rahm, E. and Do, H. H. (2000) 'Data Cleaning: Problems and Current Approaches', *Bulletin of the Technical Committee on Data Engineering*, 23(4), pp. 3–13. doi: 10.1145/1317331.1317341.

Rajabifard, A. (2010) *Critical issues in global geographic information management- with a detailed focused on Data Integration and Interoperability of Systems and Data*, *The*

2nd Prepatory Meeting of the Proposed UN Committee on Global Geographic Information Management. New York. Available at: http://www.csdila.unimelb.edu.au/people/abbas-rajabifard.html (Accessed: 31 August 2018).

Repa, V. (2011) ‘Business Process Modelling Notation from the Methodical Perspective’, in Cezon, M. and Wolfsthal, Y. (eds) *Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, pp. 160–171. doi: https://doi.org/10.1007/978-3-642-22760-8_18.

Rezgui, Y., Boddy, S., Wetherill, M. and Cooper, G. (2011) ‘Past, present and future of information and knowledge sharing in the construction industry: Towards semantic service-based e-construction?’, *Computer-Aided Design*. Elsevier, 43(5), pp. 502–515. doi: 10.1016/J.CAD.2009.06.005.

Riemer, F. J., Lapan, S. D. and Quartaroli, M. T. (2012) *Qualitative research: an introduction to methods and designs*. Edited by S. D. Lapan, M. T. Quartaroli, and J. R. Frances. San Francisco: Jossey-Bass.

Rietveld, L. and Hoekstra, R. (2014) ‘YASGUI: Feeling the Pulse of Linked Data’, in Janowicz, K., Schlobach, S., Lambrix, P., and Hyvönen, E. (eds) *Knowledge Engineering and Knowledge Management. EKAW 2014*. Linköping: Springer, Cham, pp. 441–452. doi: https://doi.org/10.1007/978-3-319-13704-9_34.

Robert Stackowiak, Venu Mantha, Alan Manewitz, A. L. (2015) *Improving Oil & Gas Performance with Big Data*. Available at: http://www.oracle.com/us/technologies/big-data/big-data-oil-gas-2515144.pdf (Accessed: 31 August 2018).

Roman, D., Nikolov, N., Putlier, A., Sukhobok, D., Elvesæter, B., Berre, A., Ye, X., Dimitrov, M., Simov, A., Zarev, M., Moynihan, R., Roberts, B., Berlocher, I., Kim, S., Lee, T., Smith, A. and Heath, T. (2017) ‘DataGraft: One-stop-shop for open data management’, *Semantic Web*, pp. 1–19. doi: 10.3233/SW-170263.

Rong, C. (2009) ‘An Industrial Cloud: Integrated Operations in Oil and Gas in the Norwegian Continental Shelf’, in Jaatun, M. G., Zhao, G., and Rong, C. (eds) *IEEE International Conference on Cloud Computing*. Berlin: Springer, pp. 19–23. doi: 10.1007/978-3-642-10665-1_3.

Royal Academy of Engineering (2013) *Decommissioning in the North Sea: A report of a workshop held to discuss the decommissioning of oil and gas platforms in the North Sea*. Available at: https://www.raeng.org.uk/publications/reports/decommissioning-in-the-north-sea (Accessed: 16 August 2017).

RSA and ZWS (2015) *The RSA Great Recovery & Zero Waste Scotland Programme: North Sea Oil and Gas Rig Decommissioning & Re-use Opportunity Report*. Available at: https://www.thersa.org/globalassets/pdfs/reports/rsa-great-recovery---north-sea-oil-and-gas-report.pdf (Accessed: 6 February 2018).

Rubin, V. L. (2014) ‘Veracity roadmap: Is big data objective, truthful and credible?’, *Advances in Classification Research Online*, 24, pp. 4–15. doi: 10.7152/acro.v24i1.14671.

Rusu, O., Halcu, I., Grigoriu, O., Neculoiu, G., Sandulescu, V., Marinescu, M. and Marinescu, V. (2013) ‘Converting unstructured and semi-structured data into knowledge', in *11th RoEduNet International Conference*. Sinaia, pp. 1–4. doi: 10.1109/RoEduNet.2013.6511736.

Ryu, S. (2013) *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie or Die*. John Wiley & Sons, Inc.

Saito, T. and Rehmsmmeier, M. (2017) 'Precrec: fast and accurate precision–recall and ROC curve calculations in R', *Bioinformatics*, 33(1), pp. 145–147. doi: 10.1093/bioinformatics/btw570.

Sarrab, M. and Rehman, O. M. H. (2014) 'Empirical study of open source software selection for adoption, based on software quality characteristics', *Advances in Engineering Software*. Elsevier, 69, pp. 1–11. doi: 10.1016/J.ADVENGSOFT.2013.12.001.

Schwalbe, K. (2011) *Information Technology: Project Management*. Cengage Learning. Available at: https://fisnikd.files.wordpress.com/2016/04/ebook-kathy-schwalbe.pdf (Accessed: 6 February 2018).

Shao, W., Zhang, H., Liu, J., Yang, G., Chen, X., Yang, Z. and Huang, H. (2016) 'Data Integration and its Application in the Sponge City Construction of China', in *12th International Conference on Hydroinformatics - Smart Water for the Future*, pp. 779–789. doi: https://doi.org/10.1016/j.proeng.2016.07.583.

Shayeganfar, F., Mahdavi, A., Suter, G. and Anjomshoaa, A. (2009) 'Implementation of an IFD library using semantic web technologies: A case study', *Ework and Ebusiness in Architecture, Engineering and Construction*, pp. 539–544.

Siems, G. L. (2016) 'Decommissioning Process Optimization Methodology', in *Offshore Technology Conference*. Houston. doi: 10.4043/26867-MS.

Simon, H. (1972) 'Theories of Bounded Rationality', in *Decision and Organization*. North Holland Publishing Co., pp. 161–176.

Šindelář, R. and Novák, P. (2016) 'Ontology-Based Simulation Design and Integration', in Biffl, S. and Sabou, M. (eds) *Semantic Web Technologies for Intelligent Engineering Applications*. Springer, Cham, pp. 257–277. doi: doi.org/10.1007/978-3-319-41490-4_10.

Sing, T., Sander, O., Beerenwinkel, N. and Lengauer, T. (2005) 'ROCR: Visualizing classifier performance in R', *Bioinformatics*, 21(20), pp. 3940–3941. doi: 10.1093/bioinformatics/bti623.

Soibelman, L., Wu, J., Caldas, C., Brilakis, I. and Lin, K.-Y. (2008) 'Management and analysis of unstructured construction data types', *Advanced Engineering Informatics*. Elsevier, 22(1), pp. 15–27. doi: 10.1016/J.AEI.2007.08.011.

Soleimanifar, M. (2016) *Integrated Project Management Framework for a Pipe Spool Fabrication Shop*. University of Alberta. doi: doi:10.7939/R36M33G31.

Son, H., Bosché, F. and Kim, C. (2015) 'As-built data acquisition and its use in production monitoring and automated layout of civil infrastructure: A survey', *Advanced Engineering Informatics*. Elsevier Ltd, 29(2), pp. 172–183. doi: 10.1016/j.aei.2015.01.009.

Sork, T. J. (1997) ‘Workshop Planning’, *New Directions for Adult and Continuing Education*, 1997(76), pp. 5–17. doi: 10.1002/ace.7601.

Spyns, P., Meersman, R. and Jarrar, M. (2002) ‘Data modelling versus ontology engineering’, *ACM SIGMOD Record*, 31(4), p. 12. doi: 10.1145/637411.637413.

Stacey, A., Sharp, J. V and Birkinshaw, M. (2008) *Life Extension Issues for Ageing Offshore Installations*, *Proceedings of 27th International Conference on Offshore Mechanics and Arctic Engineering*. doi: doi:10.1115/OMAE2008-57411.

Standards Norway (1996) *Piping Fabrication, Installation, Flushing and Testing*. Standards Norway. Available at: http://www.ccc.no/publish_files/Norsok_L-004-CR_Ror_og_trykk_Engelsk.pdf (Accessed: 6 February 2018).

Steele, K. and Stefánsson, H. O. (2016) *Decision Theory*, *The Stanford Encyclopedia of Philosophy*. Available at: https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=decision-theory (Accessed: 17 February 2018).

Stidolph, N. (2015) *What the Oil and Gas Industry Can Tell You About Documents and Data*. San Diego.

Studer, R., Benjamins, V. R. and Fensel, D. (1998) ‘Knowledge engineering: Principles and methods’, *Data & Knowledge Engineering*, 25(1–2), pp. 161–197. doi: 10.1016/S0169-023X(97)00056-6.

Sun, M. and Aouad, G. (1999) ‘Control Mechanism for Information Sharing in an Integrated Construction Environment’, in *Proceeding of The 2nd International Conference on Concurrent Engineering in Construction*, pp. 1–10.

TechInvestLab.ru (2013) *.15926 Editor*. Available at: http://techinvestlab.ru/files/V4/dot15926Editor14_Vol4_PatternsAndMapping.pdf (Accessed: 13 February 2018).

The W3C SPARQL Working Group (2013) *SPARQL 1.1 Overview*, *W3C Recommendation*. Available at: https://www.w3.org/TR/sparql11-overview/ (Accessed: 6 February 2018).

Tolk, A. and Diallo, S. Y. (2005) ‘Model-based data engineering for web services’, *IEEE Internet Computing*, 9(4), pp. 65–70. doi: 10.1109/MIC.2005.87.

Tserng, H. P., Yin, S. Y. L., Dzeng, R. J., Wou, B., Tsai, M. D. and Chen, W. Y. (2009) ‘A study of ontology-based risk management framework of construction projects through project life cycle’, *Automation in Construction*. Elsevier, 18(7), pp. 994–1008. doi: 10.1016/J.AUTCON.2009.05.005.

UK Gov (2008) ‘Energy Act 2008’, 32(4). Available at: http://www.legislation.gov.uk/ukpga/2008/32/pdfs/ukpga_20080032_en.pdf%5Cnhttp://www.legislation.gov.uk/ukpga/2008/32/contents.

Vasilev, M. V. (2018) *Decom Waste Trader Project using Linked data API*. Heriot-Watt University.

Vestola, M. (2010) ‘A Comparison of Nine Basic Techniques for Requirements Prioritization’, *Helsinki University of Technology*, pp. 1–8. Available at:

https://pdfs.semanticscholar.org/2ff7/d2c0fe7da8fd3dcd22e635335a33b675fcd2.pdf.

Volk, R., Stengel, J. and Schultmann, F. (2014) ‘Building Information Modeling (BIM) for existing buildings - Literature review and future needs’, *Automation in Construction*. Elsevier B.V., 38, pp. 109–127. doi: 10.1016/j.autcon.2013.10.023.

Walsh, C. (2016) *Data and Analytics: Open Source Data Integration Tool Comparison*. Available at: https://www.excella.com/wp-content/uploads/2016/03/Open-Source-DI-Tool-Comparison_March2016.pdf (Accessed: 6 February 2018).

Wang, X., Li, H., Wong, J. and Li, H. (2014) ‘An integration framework of advanced technologies for productivity improvement for LNG mega-project’, *Journal of Information Technology in Construction*, 19(22), pp. 360–382. doi: http://www.itcon.org/2014/22.

Watson, H. J. (2009) ‘Tutorial: Business Intelligence – Past, Present, and Future’, in *Communications of the Association for Information Systems*, pp. 487–510.

WellMet2050 (2010) *Conserving our Metal Energy*. Available at: http://www.lcmp.eng.cam.ac.uk/wp-content/uploads/WellMet2050-Conserving-our-metal-energy-Sept-2010-Web.pdf (Accessed: 23 February 2018).

Westin, S. and Sein, M. K. (2014) ‘Improving Data Quality in Construction Engineering Projects : An Action Design Research Approach’, *Journal of Management in Engineering*, 30(3), pp. 1–11. doi: 10.1061/(ASCE)ME.1943-5479.0000202.

White, M. (2018) ‘Big Data vs . Data Warehousing’, *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 3(5), pp. 381–383. Available at: http://ijsrcseit.com/paper/CSEIT183595.pdf.

Wiesner, A., Morbach, J. and Marquardt, W. (2011) ‘Information integration in chemical process engineering based on semantic technologies’, *Computers and Chemical Engineering*. Elsevier Ltd, 35(4), pp. 692–708. doi: 10.1016/j.compchemeng.2010.12.003.

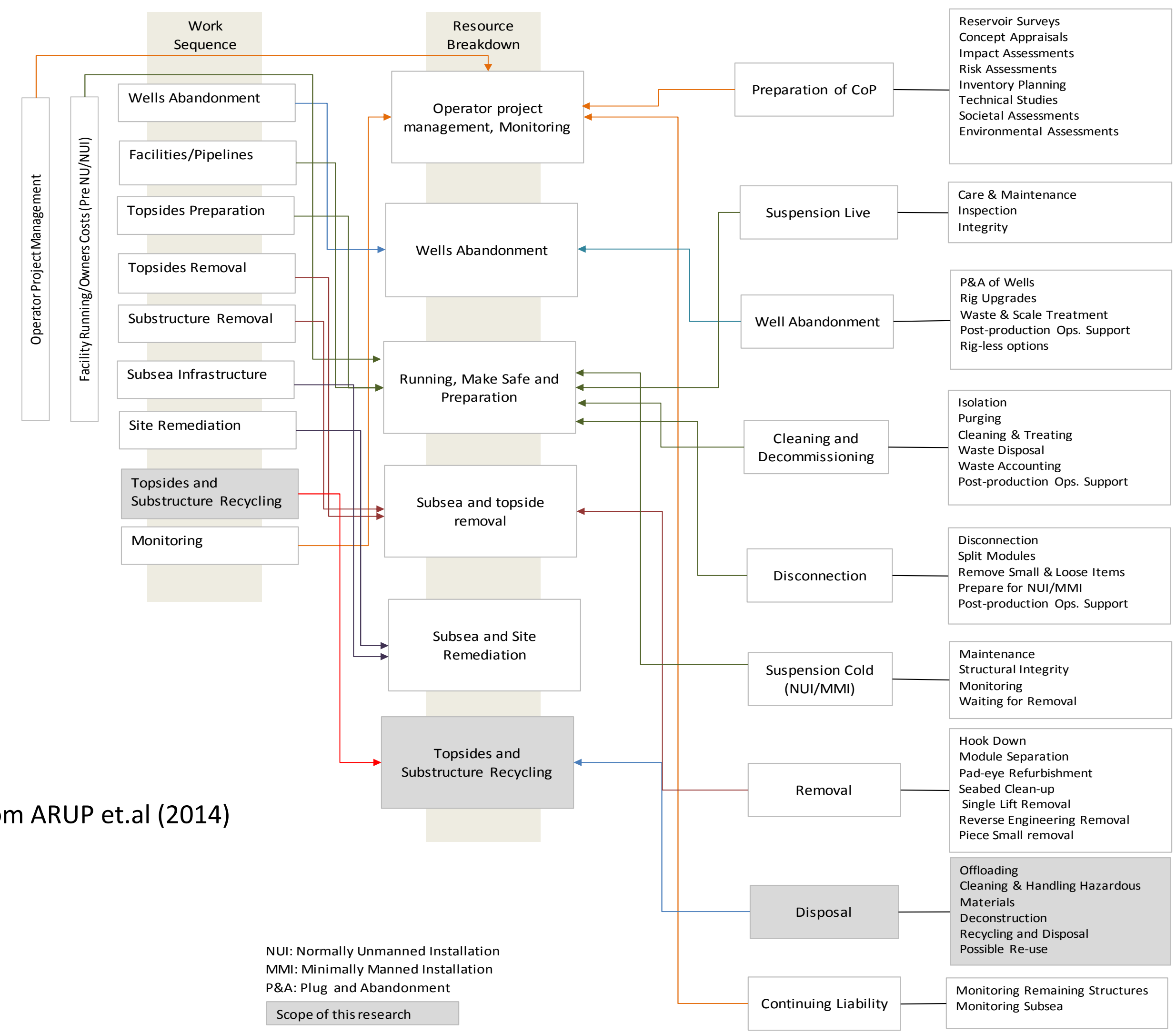
Xu, L. and Embley, D. W. (2004) ‘Combining the Best of Global-as-View and Local-as-View for Data Integration’, *Proceedings of the 3rd International Conference ISTA’2004: Information Systems Technology and its Applications*, (December 2002), pp. 123–135. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.6.9062.

Yang, R. (2009) *Process Plant Lifecycle Information Management*. Bloomington, IN: iUniverse.

Yurchyshyna, A. and Zarli, A. (2009) ‘An ontology-based approach for formalisation and semantic organisation of conformance requirements in construction’, *Automation in Construction*. Elsevier B.V., 18(8), pp. 1084–1098. doi: 10.1016/j.autcon.2009.07.008.

Appendices

Appendix 1: Offshore Decommissioning Work Breakdown Structure



Adapted from ARUP et.al (2014)

Appendix 2: Output from workshops

Table A.1: Summary of attended conferences and workshops

S/N	Conference/Workshop	Year	Location	Theme	Relevant Topics	Important Summaries	Research Nexus/Comments
1	Decommissioning and Wreck removal Workshop	2015	Glasgow, UK	Best practice for managing Man Made Structures on the Seafloor over their life-cycle, using an evidence -based approach	Environmental Considerations for wreck removal	OSPAR Regulation requires all topsides of offshore oil and gas installations to be returned to land and managed in an environmentally friendly way. Given this, reuse is more environmentally friendly than recycle. Consequently reuse of decommissioned items should be encouraged.	This informed the focus on reuse of offshore decommissioning waste items as a case study.
					Transferring methodologies from reference industries	Established methods and processes in other industries e.g. construction and nuclear decommissioning sectors should be explored for offshore decommissioning.	Steel reuse in the construction industry is a motivation for exploring steel reuse in offshore decommissioning for this study.
					Data Science and innovation opportunities	Advances in data science should be employed to manage big and complex lifecycle data to achieve higher efficiencies in the decommissioning process.	This is a basis for exploring appropriate data integration technologies for the use cases in this study.
2	Offshore Decommissioning Conference	2015	St Andrews, UK	Decommissioning -Taking the Initiative	Re-Use of Offshore Oil and Gas Infrastructure	Early re-use assessment is an important step in the reuse of oil and gas infrastructure. However, this is dependent on availability of documentation to determine re-sale potential.	All the available information, irrespective of format or form, will be required in making the best judgement on the potential for reuse of an item. Consequently, technologies that can integrate heterogenous data are a research consideration.
					Taxpayers Interest in Public Sector Projects	There is an increased interest in decommissioning compared to other sectors because government has decided to procure assurances against all the associated risks. Until government optimises its risks liability with respect to decommissioning, focus will continue to be on the sector. As a result efficient ways to mitigate the risks are very important.	This is a reason why maximising economic recovery from decommissioning is important. As a result, the direction of this research to support reuse and possibly resale of decommissioned items is in the right direction.
3	Decommissioning and Wreck removal Workshop	2016	Glasgow, UK	Influence, educate and un-lock potential – Bringing industry and science together	The benefits of a circular economy approach during decommissioning	Circular economy could add over £500 million to end of life value of assets deployed in Scottish water alone. Particularly, it is a huge opportunity for the reuse of steel infrastructure.	This is a reason why steel reuse case study was prioritised.
					Subsea Decommissioning - Challenges and Learnings	Information gathering is key to subsea decommissioning. Information in different formats is received from the client during tendering and additional information may be gathered by the contractor using different techniques where required. Data gathered from the multiple sources need to be synthesised correctly to achieve a successful decommissioning campaign.	This is a basis for consideration of data integration technologies.
					Science for Decom	At decommissioning, data available for planning are supplied by multiple stakeholders, exist in different formats, have error and missing values, etc. Consequently, there is a need to integrate them. One possible way of doing this is the use of ISO 15926.	This is a basis for considering ISO 15926 application in offshore decommissioning process data integration

Appendix 2: Output from Conference Workshops *(Continued)*

Table A.1: Summary of attended conferences and workshops

S/N	Conference/Workshop	Year	Location	Theme	Relevant Topics	Important Summaries	Research Nexus/Comments
4	Offshore Decommissioning Conference	2016	Aberdeen, UK	Reality, preparation and action in a low oil price environment	Centrica Decommissioning Projects Update	Centrica has a decommissioning project that has a target of cost neutrality . It intends to achieve this through sales and reuse of equipment from the facility.	This validates the importance attached to reuse opportunities. As a result, an additional justification for prioritizing reuse case study in this research.
					DataCo Exhibition on Data and Information Management	The company helps decommissioning team to discover critical data and identify gaps in their knowledge base. It does this by delivering dashboards for decisions and querying of integrated data. This is supposed to save time engineers spend searching for answers in data.	This validates the existence of efforts to apply Business Intelligence techniques to offshore decommissioning process.
5	Offshore Technology Conference	2016	Houston, USA	Endless Limitations	Decommissioning Process Optimisation Methodology	Decommissioning methods that could get more work done, safely and at a lower cost, are desired because traditional methods, though get the work done, are very slow. By employing the newest technologies, a new, optimised decommissioning methodology is possible.	New technology in this context refers to large vessels that can carry out single lift of topsides of offshore installations. Conversely, it can also refer to piece small removal method that allows the early removal of useful components from installation to limit the level of deterioration that could result from a normal process - due to vessel wait times, and increase reuse opportunities. In order to carry out piece small . This process will require repeated audit of the balance of items on the platform. This is the motivation for 'use case 2' in this research.
					Integrated Decommissioning – Increasing Efficiency	A more efficient method of platform decommissioning should integrate multiple services and manage them as a single project with the objective of completing the decommissioning job safely and efficiently at a fixed cost to the operator. This will mitigate the risks of the operator and efficiency is emphasised for the service contractor.	Integration of processes implies integration of data. As a result, an indication of the need for data integration techniques in offshore decommissioning.
					Smart Production Monitoring & Management Information Systems	It is important to use software technology applications that can take care of hidden semantics when handling oil and gas data because missing or erroneous details, even though small, can have serious implications. Secondly, reduction of costs through efficient processes requires software technology application and a promising example is Big Data.	This indicates the relevance of Big Data in offshore decommissioning process.

Appendix 2: Output from Conference Workshops *(Continued)*

Table A.2: Summary of unstructured interview responses on requirements for process plant data integration

S/N	Questions	Comment Summaries	Solution Requirements
1	Use of ISO 15925	✔ ISO 15926 covers the lifecycle of a process plant, so will be appropriate for integrating data at decommissioning ✔ Stakeholders and application developers seek to comply with ISO 15926 because it is a global standard ✔ Engineers are not very familiar with the technologies required for the proper use of ISO 15926 ✔ Lack of examples on the use of ISO 15926 for data integration is a challenge for potential users	✔ Appplication of ISO 15926 ✔ Tools for using ISO 15926
2	Systems architecture	✔ Solution should be able to connect to live autonomous sources, particularly relational database ✔ Solution should be able to handle multiple sources simultaneously	✔ Querying technique
3	Inventory development	✔ Need for ways to reduce the effort required for extracting data from different sources ✔ Need for ways to eliminate costly human errors inherent in manual approaches ✔ Need for ways to resolve meaning of terms ✔ Need for ways to resolve data level representation	✔ Smooth data migration ✔ Methods for handling of common data types ✔ Syntatic and semantic homogeneity
4	Data extraction	✔ Data output should be in the order desired ✔ Need for capability to query across multiple sources ✔ Data output should indicate where missing values exist	✔ Querying technique
5	Desirable attributes	✔ Solution should require minimal effort and return very accurate result ✔ Solution should result in reduced task duration ✔ The IT artifacts generated by the solution should be reusable	

S/N	Conference/Workshop	Year	Location	Participants	Industry Experience
1	Decommissioning and Wreck removal Workshop	2015	Glasgow, UK	5	10 - 20
2	Offshore Decommissioning Conference	2015	St Andrews, UK	3	15 - 20
3	Decommissioning and Wreck removal Workshop	2016	Glasgow, UK	5	10 - 20
4	Offshore Decommissioning Conference	2016	Aberdeen, UK	6	8 - 25
5	Offshore Technology Conference	2016	Houston, USA	2	20 - 30
			Total	21	

Appendix 3: Research Use Case Templates

Use Case Identifier	[1] Construction information management
Description	Project engineer retrieves the true status of any spool undergoing fabrication **source** [Ariosh (2018)]
Actors	Project engineer (primary) Engineering Department (secondary) X-ray Contractor (secondary) Leak Test Contrator (secondary)
Assumptions	Heterogenous sources and data formats The x-ray of welded joints precedes pressure tests in the work process, and reports from the test activities are used to update the project engineer's overall project status. All data can be processed in spreadsheets.
Steps	1. IN PARALLEL 1.1 Preprocess data in sources 1.2 Create an ontology of the domain 2. Create mapping files 3. REPEAT 3.1 Map the prepared data in the spreadsheets to the created ontology 3.2 Query integrated data
Variation	**#3.2** Query via GUI tools when there are no business rules **or** via SPARQL endpoint linked to triplestore when rules are required
Non-Functional Requirements	Perfomance: Time must be less than manual approach Reliability: Status of spool must be correct Reasoning: Allow application of business rules and resolve relationship between data entities
Issue	Decentralised software architecture, software artifacts need to be coupled into a single application

Figure A.1: Use case template for construction information management

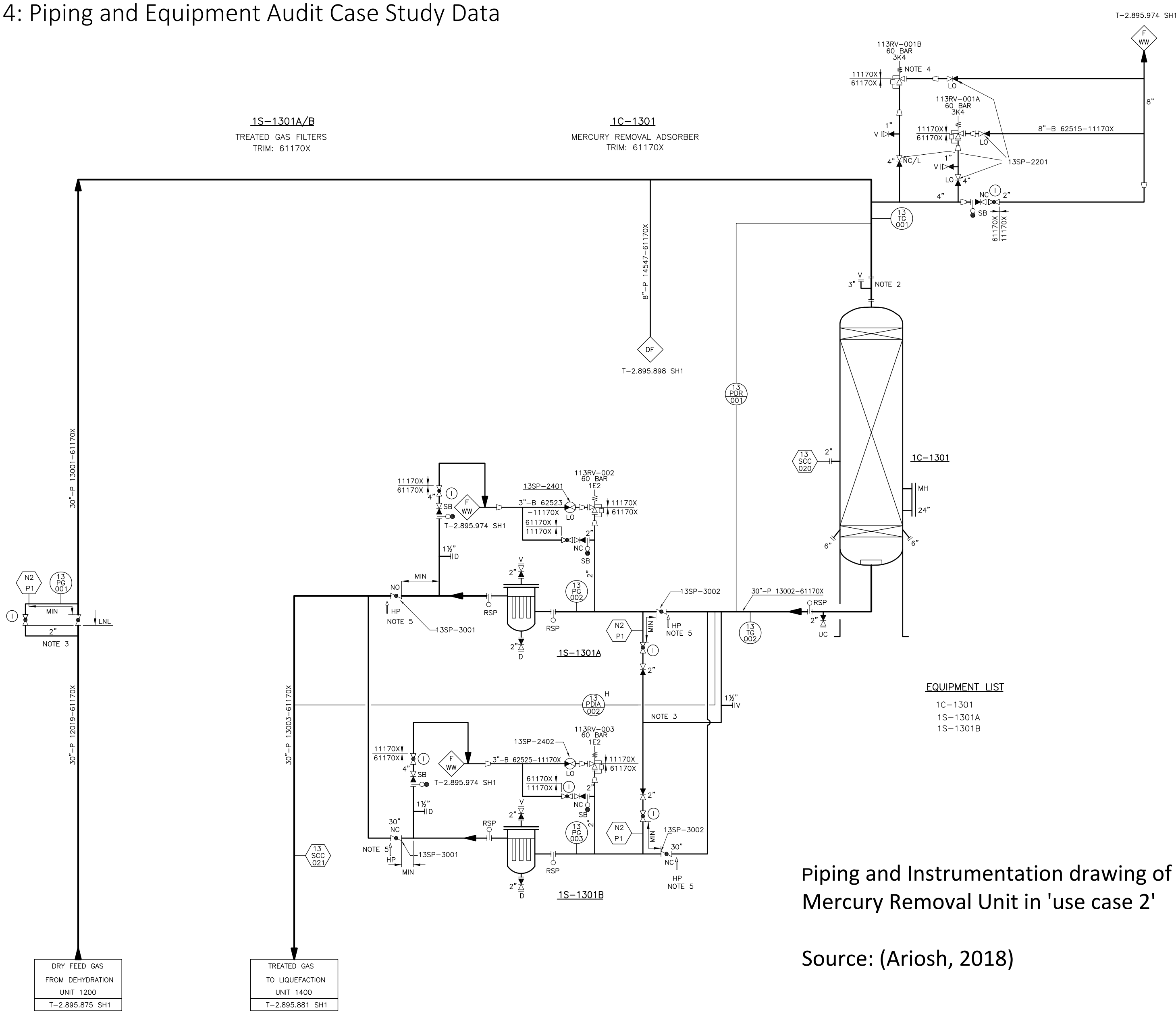
Use Case Identifier	[2] Piping and equipment audit
Description	Design engineer counts the number of a given type of line item **source** [Ariosh (2018)]
Actors	Design engineer (primary) Maintenance department (secondary) Engineering department (secondary)
Assumptions	Heterogenous sources and data formats Data can be processed in spreadsheets
Steps	1. IN PARALLEL 1.1 Preprocess data in sources 1.2 Create an ontology of the domain 2. Create mapping files 3. REPEAT 3.1 Map the prepared data in the spreadsheets to the created ontology 3.2 Query integrated data
Variation	**#3.2** Query RDF file using GUI tools **or** store data in triplestore and query via SPARQL endpoint
Non-Functional Requirements	Perfomance: Time spent must be less than that of manual approach Reliability: Count results must be correct Reasoning: Resolved relationships between data entities
Issues	Decentralised software architecture, software artifacts need to be coupled into a single application

Figure A.2: Use case template for piping and equipment audit

Use Case Identifier	[3] Steel waste assessment in offshore decommissioning
Description	Decommissioning engineer assesses steel at the end of life of a facility **sources** [Decommissiong Conferences 2015, 2016], [DNS, AMEC and ZWS (2016)]
Actors	Decommissioning engineer (primary) Maintenance department (secondary) Operations department (secondary)
Assumptions	Heterogenous sources and data formats Relational Database (RDB) needs to be queried directly Data can be processed in spreadsheets All data entities have representation in standard ontology (ISO 15926)
Steps	1. IN PARALLEL 1.1 Create an ontology of the entities needed from the RDB 1.2 Create an OBDA model for mapping data from the RDB 1.3 Preprocess data in single file sources in spreadsheets 1.4 Create ISO 15926 templates for entities in the file sources 1.5 Create mapping files for the file sources 2. Map data from single file sources to ISO 15926 templates 3. Store integrated data for the file sources in a triplestore 4. Set up virtual triplestore using the created OBDA model and ontology of entities need from the RDB 5. Query the RDB and triplestore via a SPARQL endpoint
Variation	**#5** Data is queried via a SPARQL endpoint **or** via SPARQL query GUI with access to the triplestores
Non-Functional Requirements	Perfomance: Time spent must be less than that of manual approach Reliability: Assessment information must be correct Reasoning: Resolved relationships between data entities
Issues	Decentralised software architecture, software artifacts need to be coupled into a single application

Figure A.3: Use case template for offshore decommissioning steel waste assessment

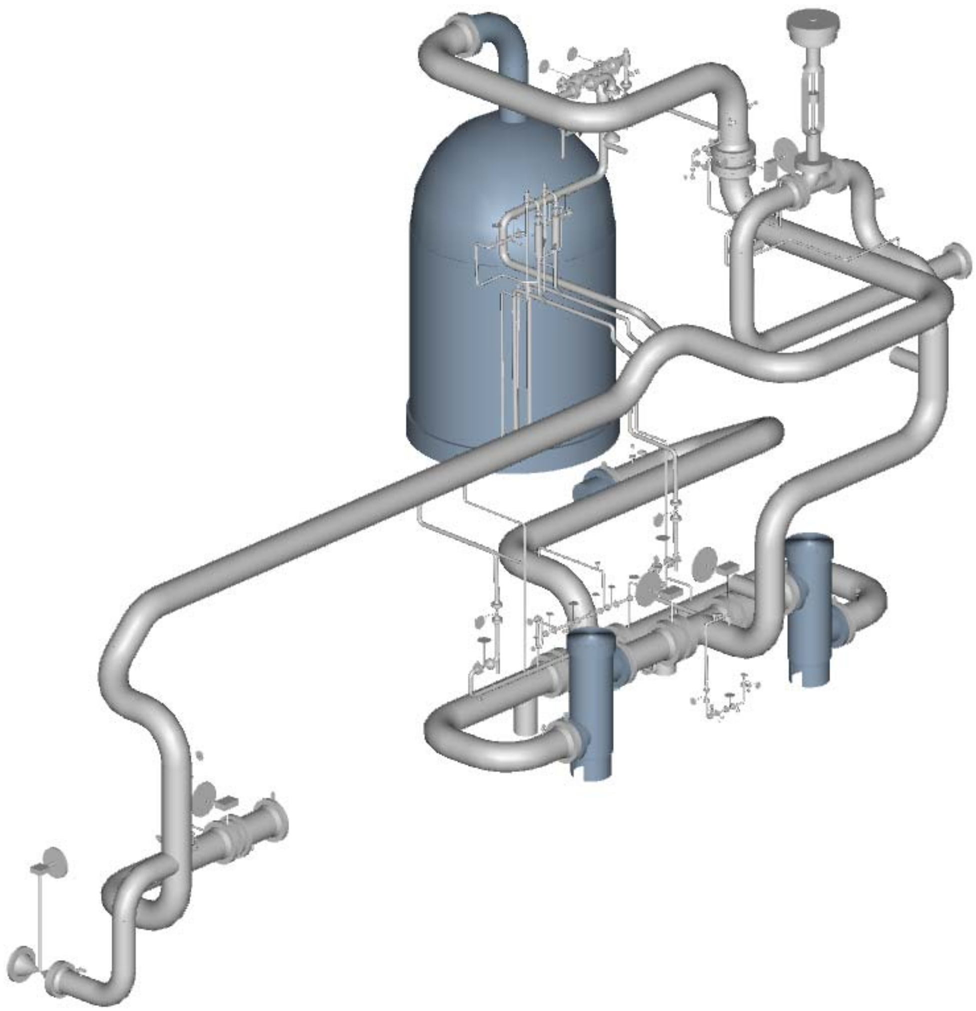
Appendix 4: Piping and Equipment Audit Case Study Data



Piping and Instrumentation drawing of Mercury Removal Unit in 'use case 2'

Source: (Ariosh, 2018)

Appendix 4: Piping and Equipment Audit Case Study Data *(Continued)*



3D model of the Mercury Removal Unit in 'use case 2'

Source: (Ariosh, 2018)

Appendix 4: Piping and Equipment Audit Case Study Data *(Continued)*

Data Source Summary

Source	Source Description	Data
Export from Maintenance Database	Maintenance records	Contact
		Equipment
		Maintenance
		StaffList
		Services
Export from Design Model	Operating parameters	Pipe
		Equipment
Export from Design Model	Piping isometrics material take off	3-B62523-11170X
		3-B62525-11170X
		8-B62515-11170X
		8-P14547-61170X
		30-P12019-61170X
		30-P13001-61170X
		30-P13002-61170X
		30-P13003-61170X

Maintenance database relational schema

StaffList	(<u>StaffID</u>, FirstName, LastName, Role, Department)
Maintenance	(<u>ID</u>, *equipmentLabel*, Service Date, *serviceID*, *Validator*, NextSchedule, Remarks)
Services	(<u>ID</u>, *ContactID*, Label, Description)
Equipment	(<u>Label</u>, Description, *ManufacturerID*, PurchaseDate, *ResponsiblePersonID*, AuthorisedUser, FacilityUnit, PhysicalCondition)
Contact	(<u>ID</u>, Company, Phone, Website)

Notes

1. <u>StaffID</u> = *Validator* = *ResponsiblePersonID*
2. Contact (<u>ID</u>)= *ContactID* = ManufacturerID
3. <u>Underlined</u> words are primary keys
4. *Italicised* words are foreign keys

Appendix 4: Piping and Equipment Audit Case Study Data *(Continued)*

		Contact	
ID	Company	Phone	Website
C0001	MACAW Engineering Ltd	+441912154010	www.macawengineering.com
C0002	Brown Corrosion Services Inc	+18323270965	browncorrosion.com
C0003	Corrosion Doctors		corrosion-doctors.org
C0004	Sea Tech Offshore		seatechoffshore.com
C0005	Mountain High Investigations	+17073670802	www.mountainhighinvestigations.com
C0006	Visions Enterprise	+17804858500	www.visions-enterprise.com
C0007	Silverwing Middle East LLC	+441792798711	www.silverwingme.com
C0008	Tubecare International	+97444275589	www.tubecare.net
C0009	ETS Risk Management	+18774552935	www.ets-riskmanagement.com
C0010	BRMS	+18327329105	www.bopriskmitigation.com
C0011	Ariosh Ltd	+23412705825	www.ariosh.com
C0012	HW Fairway	+18888134654	www.hwfairway.com
C0013	C-Tecnics	+441224608000	www.c-tecnics.com
C0014	PEC Laser	+18557322887	www.peclaser.com
C0015	Digital Surveys Ltd	+441224900133	www.digitalsurveys.co.uk
C0016	TEKDrilling Solutions LLC	+2812568181	www.tekdrilling.com
C0017	Measurement Technology Northwest	+12066341308	lci.mtnw-usa.com
C0018	Dimension Energy Services	+18068263463	dimensionenergyservices.com
C0019	JSB Engineering Consultants		
C0020	CDM Automation	+19722355624	www.cdmautomation.com
C0021	SonaCorps Inc	+17134741555	www.sonacorps-ndt.com
C0022	Structural Rig Surveys	+441349856888	www.structural-rig-surveys.com
C0023	Clover Tool Company	+441224780700	clovertool.com
C0024	Houston Boom Truck	+17133640156	www.houstonboomtruck.com
C0025	SGS	+41227399111	www.sgs.com/en
C0026	Sea Tech Offshore		seatechoffshore.com
C0027	Drill Quip	+61746343500	www.drillquip.net
C0028	IMX Ocean ConsultantsIMX Ocean Consultants	+912227711006	www.imxconsult.com
C0029	Paradigm Consultants	+17136866771	paradigmconsultants.com
C0030	Northwest Technical Solutions	+18324635081	www.northwest-technical.com
C0031	Serious Engineering	+448434874088	www.seriousengineering.co.uk
C0032	Southern Metal Fabricators	+18009891330	www.southernmetalfab.com
C0033	WorthFab LLC		www.worthfab.com
C0034	United Fabrication	+17806572509	www.unitedfab.ca
C0035	Eversendai Offshore	+97144564815	www.eversendaioffshore.com
C0036	ICON Engineering	+61863135500	www.iconeng.com.au
C0037	Boswell Oil Gas Services	+2812527171	www.boswelloilandgas.com
C0038	LDD Ltd	+441209861930	www.lddrill.com
C0039	Daldoss	+390461518647	www.daldoss.com
C0040	Mpower	+6128788 4600	www.mpower.com.au
C0041	Firbimatic	+390516814189	www.metalcleaning-firbimatic.com
C0042	Chemical Engineering Services		www.spsb.s5.com
C0043	DP Fuel Tank Services	+441737767524	www.dptanks.co.uk
C0044	Octane	+441132012460	www.octane.uk.com
C0045	Vac Qua	+13307740211	www.vacqua-llc.com
C0046	Ecoserv	+13379844445	www.ecoserv.net
C0047	EquipmentAuction.com LLC	+13303663025	www.equipmentauction.com
C0048	DP Fuel Tank Services	+441737767524	DP Fuel Tank Services
C0049	Resmar Ltd	+4408458033399	www.resmar.co.uk
C0050	3Di Equipment	+441606738766	www.3diequipment.com
C0051	RPM Fuels and Tanks	+441473787787	www.rpm-fuels.co.uk
C0052	United Filtration Systems	+15868025561	unitedfiltration.com
C0053	Micronics Inc	+8664813694	www.micronicsinc.com
C0054	Buffalo Flange Inc	+17139439600	www.buffaloflange.com
C0055	Esbjerg Oiltool AS	+457515 6143	www.estool.dk
C0056	Ezeflow Inc	+17246583711	ezeflow.com
C0057	Vestek Industries	+1713-242-7700	www.vestekindustries.com
C0058	Southwest Oilfield Products Inc	+17136757541	www.swoil.com
C0059	Apex	+6564835350	www.apex-chemicals.com

Appendix 4: Piping and Equipment Audit Case Study Data *(Continued)*

Equipment

Label	Description	ManufacturerID	PurchaseDate	ResponsiblePersonID	AuthorisedUser	FacilityUnit	PhysicalCondition
1-C-1301	Mercury Removal Adsorber	C0058	4-Jan-10	AR010	AR006	TR1-PIPES-U1300-P	Corroded vessel
1S-1301A	Treated Gas Filter	C0057	30-Nov-09	AR010	AR011	TR1-PIPES-U1300-P	Corroded pipework
1S-1301B	Treated Gas Filter	C0057	31-Aug-09	AR010	AR011	TR1-PIPES-U1300-P	Leaks

Maintenance

ID	equipmentLabel	Service Date	serviceID	Validator	NextSchedule	Remarks
M001	1S-1301A	31-Jan-10	S012	AR002		Completed
M002	1S-1301A	3-Feb-10	S007	AR002		Completed
M003	1S-1301A	4-Mar-10	S008	AR002		Completed
M004	1-C-1301	8-Feb-13	S035	AR004	8-Mar-14	Completed
M005	1-C-1301	14-Mar-13	S062	AR004		Completed
M006	1-C-1301	20-Mar-13	S044	AR006		Completed
M007	1S-1301B	21-Mar-13	S049	AR006		Completed

StaffList

Staff ID	FirstName	LastName	Role	Department
AR001	Kenny	Adeyemi	Project Manager	Projects
AR002	Yusuf	Alege	Engineering Manager	Engineering
AR003	Bako	Adegboyega	General Manager	Procurement
AR004	Posi	Atoyebi	Acquisition Manager	Procurement
AR005	Doyin	Olaniran	Development Manager	Procurement
AR006	Yomi	Obaoye	Principal Engineer	Engineering
AR007	Ola	Oshe	Chief Technical officer	
AR008	Ponle	Oshe	Development Manager	Procurement
AR009	Niyi	Adebayo	Logistics Manager	Procurement
AR010	Gbenga	Adeshina	Construction Manager	Construction
AR011	Damilola	Adeyemi	Piping Engineer	Engineering
AR012	Ronald	Makanjuola	Structural Engineer	Engineering

Appendix 4: Piping and Equipment Audit Case Study Data *(Continued)*

Services

ContactID	Label	Description
C0001	Forensic Investigation	Analysis of failed equipment
C0002	Forensic Investigation	Analysis of failed equipment
C0003	Forensic Investigation	Analysis of failed equipment
C0004	Forensic Investigation	Analysis of failed equipment
C0005	Forensic Investigation	Analysis of failed equipment
C0001	Risk-Based Inspection	Analysis of fixed equipment for corrosion and stress cracking
C0006	Risk-Based Inspection	Analysis of fixed equipment for corrosion and stress cracking
C0007	Risk-Based Inspection	Analysis of fixed equipment for corrosion and stress cracking
C0008	Risk-Based Inspection	Analysis of fixed equipment for corrosion and stress cracking
C0009	Risk-Based Inspection	Analysis of fixed equipment for corrosion and stress cracking
C0010	Risk-Based Inspection	Analysis of fixed equipment for corrosion and stress cracking
C0001	3D Laser Scanning	3D data collection on equipment
C0011	3D Laser Scanning	3D data collection on equipment
C0012	3D Laser Scanning	3D data collection on equipment
C0013	3D Laser Scanning	3D data collection on equipment
C0014	3D Laser Scanning	3D data collection on equipment
C0015	3D Laser Scanning	3D data collection on equipment
C0001	Dimension Control	Managing the dimensions of fabrications for installation
C0011	Dimension Control	Managing the dimensions of fabrications for installation
C0016	Dimension Control	Managing the dimensions of fabrications for installation
C0017	Dimension Control	Managing the dimensions of fabrications for installation
C0018	Dimension Control	Managing the dimensions of fabrications for installation
C0019	Dimension Control	Managing the dimensions of fabrications for installation
C0020	Dimension Control	Managing the dimensions of fabrications for installation
C0001	Non-Destructive Testing	Detection of defects and irregularities in equipment
C0021	Non-Destructive Testing	Detection of defects and irregularities in equipment
C0022	Non-Destructive Testing	Detection of defects and irregularities in equipment
C0023	Non-Destructive Testing	Detection of defects and irregularities in equipment
C0024	Non-Destructive Testing	Detection of defects and irregularities in equipment
C0025	Non-Destructive Testing	Detection of defects and irregularities in equipment
C0026	Technical Inspection	Audit of equipment for process improvement
C0026	Technical Inspection	Audit of equipment for process improvement
C0027	Technical Inspection	Audit of equipment for process improvement
C0028	Technical Inspection	Audit of equipment for process improvement
C0029	Technical Inspection	Audit of equipment for process improvement
C0030	Technical Inspection	Audit of equipment for process improvement
C0011	Fabrication	Pipe welding, threading, painting and coupling
C0031	Fabrication	Pipe welding, threading, painting and coupling
C0032	Fabrication	Pipe welding, threading, painting and coupling
C0033	Fabrication	Pipe welding, threading, painting and coupling
C0034	Fabrication	Pipe welding, threading, painting and coupling
C0035	Fabrication	Pipe welding, threading, painting and coupling
C0011	Installation	Positioning and connection of fabricated pipework and equipment
C0036	Installation	Positioning and connection of fabricated pipework and equipment
C0037	Installation	Positioning and connection of fabricated pipework and equipment
C0038	Installation	Positioning and connection of fabricated pipework and equipment
C0039	Installation	Positioning and connection of fabricated pipework and equipment
C0040	Installation	Positioning and connection of fabricated pipework and equipment
C0041	Cleaning	Cleaning of equipment
C0042	Cleaning	Cleaning of equipment
C0043	Cleaning	Cleaning of equipment
C0044	Cleaning	Cleaning of equipment
C0045	Cleaning	Cleaning of equipment
C0046	Waste Management	Disposal of decomissioned equiment and pipework
C0047	Waste Management	Disposal of decomissioned equiment and pipework
C0048	Waste Management	Disposal of decomissioned equiment and pipework
C0049	Waste Management	Disposal of decomissioned equiment and pipework
C0050	Waste Management	Disposal of decomissioned equiment and pipework
C0051	Equipment Supply	Supplier of adsorber and filters
C0052	Equipment Supply	Supplier of adsorber and filters
C0053	Equipment Supply	Supplier of adsorber and filters

Appendix 4: Piping and Equipment Audit Case Study Data *(Continued)*

Pipe

LINE NO.	OWNER	TEMPERATURE	PRESSURE	FLUID	PHASE	FDENSITY	DPRESSURE	TPRESSURE	DTEMPMIN	DTEMPMAX	INSP.CLASS	PAINTCODE	TESTMEDIA	PFD
30-P13003-61170X	TR1-PIPES-U1300-P	23degC	4950000pascal	NG	V	43	6000000pascal	9170000pascal	-20degC	80degC	Class4	A	H	T-2.895.879-SH1-U
30-P13001-61170X	TR1-PIPES-U1300-P	23degC	5050000pascal	NG	V	43.9	6000000pascal	9170000pascal	-20degC	80degC	Class4	A	H	T-2.895.879-SH1-U
30-P13002-61170X	TR1-PIPES-U1300-P	23degC	5000000pascal	NG	V	43.4	6000000pasca	9170000pascal	-20degC	80degC	Class4	A	H	T-2.895.879-SH1-U
30-P12019-61170X	TR1-PIPES-U1200-P	23degC	5060000pascal	NG	V	44	6000000pascal	9170000pascal	-20degC	80degC	Class4	A	H	T-2.895.875-SH1-U
8-P14547-61170X	TR1-PIPES-U1300-P	-57.3degC	4940000pascal	CONDENSATE	L	434.46								T-2.895.898-SH1-T
8-B62515-11170X	TR1-PIPES-U1300-B	96.7degC	750000pascal	HC			1040000pascal	1140000pascal	-20degC	80degC	Class1	A	P	T-2.895.879-SH1-U
3-B62525-11170X	TR1-PIPES-U1300-B	-100000degC	0pascal											T-2.895.879-SH1-U
3-B62523-11170X	TR1-PIPES-U1300-B	-100000degC	0pascal											T-2.895.879-SH1-U

Equipment

EQUIPMENT NO.	OWNER	TEMPERATURE	PRESSURE	FLUID	PHASE	FDENSITY	DPRESSURE	TPRESSURE	DTEMPMIN	DTEMPMAX	INSP.CLASS	PAINTCODE	TESTMEDIA	PFD
1-C-1301	TR1-PIPES-U1300-P	23degC	5050000pascal	NG	V	43.9	6000000pascal	9170000pascal	-20degC	80degC	Class4	A	H	T-2.895.879-SH1-U
1S-1301A	TR1-PIPES-U1300-P	23degC	5000000pascal	NG	V	43.4	6000000pasca	9170000pascal	-20degC	80degC	Class4	A	H	T-2.895.879-SH1-U
1S-1301B	TR1-PIPES-U1300-P	23degC	5000000pascal	NG	V	43.4	6000000pasca	9170000pascal	-20degC	80degC	Class4	A	H	T-2.895.879-SH1-U

Appendix 4: Piping and Equipment Audit Case Study Data *(Continued)*

3-B62523-11170X

Type	Detail	Comment	Standard	Material	N.S. (INS)	Item Code	Length	Quantity
ELBO	ELBOW 45 DEG. BW SCH 40		ASME B16.9	ASTM A420 GR. WPL6	4	76.30.38.246.1		1
ELBO	ELBOW 45 DEG. BW SCH 40		ASME B16.9	ASTM A420 GR. WPL6	3	76.30.38.228.1		1
ELBO	ELBOW 45 DEG. BW SCH 40		ASME B16.9	ASTM A420 GR. WPL6	2	76.30.38.208.1		1
ELBO	ELBOW LR90 BW SCH 40		ASME B16.9	ASTM A420 GR. WPL6	4	76.30.39.246.1		6
ELBO	ELBOW LR90 BW SCH 40		ASME B16.9	ASTM A420 GR. WPL6	3	76.30.39.228.1		1
ELBO	ELBOW LR90 BW SCH 40		ASME B16.9	ASTM A420 GR. WPL6	2	76.30.39.208.1		4
FLAN	FLANGE 150# RFWN SCH 40		ASME B16.5	ASTM A350-LF2	3	76.70.82.110.1		2
FLAN	FLANGE 150# RFWN SCH 40		ASME B16.5	ASTM A350-LF2	2	76.70.82.080.1		1
FLAN	FLANGE 600# RFWN SCH 40		ASME B16.5	ASTM A105	4	76.62.86.080.1		1
FLAN	FLANGE 600# RFWN SCH 40		ASME B16.5	ASTM A105	2	76.62.86.076.1		1
REDU	REDU. ECC. BW SCH 40		ASME B16.9	ASTM A420 GR. WPL6	3X2	76.31.70.235.1		1
TUBI	PIPE SMLS SCH 40			ASTM A333 GR.6	4	74.13.17.075.1	15m	
TUBI	PIPE SMLS SCH 40			ASTM A333 GR.6	3	74.13.17.067.1	2.5m	
TUBI	PIPE SMLS SCH 40			ASTM A333 GR.6	2	74.13.17.059.1	3.1m	
VALV	BALL VALVE 150# RF	WITH EXTENDED LEVER OPERATOR FULL BORE, ASTM A352, A350 BODY WITH AISI 316 (L) TRIM			3	70.00.00.000.1		1

Appendix 4: Piping and Equipment Audit Case Study Data *(Continued)*

3-B62525-11170X

Type	Detail	Comment	Standard	Material	N.S. (INS)	Item Code	Length	Quantity
ELBO	ELBOW 45 DEG. BW SCH 40		ASME B16.9	ASTM A420 GR. WPL6	4	76.30.38.246.1		1
ELBO	ELBOW 45 DEG. BW SCH 40		ASME B16.9	ASTM A420 GR. WPL6	3	76.30.38.228.1		1
ELBO	ELBOW 45 DEG. BW SCH 40		ASME B16.9	ASTM A420 GR. WPL6	2	76.30.38.208.1		1
ELBO	ELBOW LR90 BW SCH 40		ASME B16.9	ASTM A420 GR. WPL6	4	76.30.39.246.1		6
ELBO	ELBOW LR90 BW SCH 40		ASME B16.9	ASTM A420 GR. WPL6	3	76.30.39.228.1		1
ELBO	ELBOW LR90 BW SCH 40		ASME B16.9	ASTM A420 GR. WPL6	2	76.30.39.208.1		4
FLAN	FLANGE 150# RFWN SCH 40		ASME B16.5	ASTM A350-LF2	3	76.70.82.112.1		2
FLAN	FLANGE 150# RFWN SCH 40		ASME B16.5	ASTM A350-LF2	2	76.70.82.110.1		1
FLAN	FLANGE 600# RFWN SCH 40		ASME B16.5	ASTM A105	4	76.62.86.080.1		1
FLAN	FLANGE 600# RFWN SCH 40		ASME B16.5	ASTM A105	2	76.62.86.076.1		1
REDU	REDU. ECC. BW SCH 40		ASME B16.9	ASTM A420 GR. WPL6	3X2	76.31.70.235.1		1
TUBI	PIPE SMLS SCH 40			ASTM A333 GR.6	4	74.13.17.075.1	15.9m	
TUBI	PIPE SMLS SCH 40			ASTM A333 GR.6	3	74.13.17.067.1	2.4m	
TUBI	PIPE SMLS SCH 40			ASTM A333 GR.6	2	74.13.17.059.1	3.1m	
VALV	BALL VALVE 150# RF	WITH EXTENDED LEVER OPERATOR FULL BORE, ASTM A352, A350 BODY WITH AISI 316 (L) TRIM			3	70.00.00.000.1		1

8-B62515-11170X

Type	Detail	Comment	Standard	Material	N.S. (INS)	Item Code	Length	Quantity
ELBO	ELBOW 45 DEG. BW SCH 30		ASME B16.9	ASTM A420 GR. WPL6	8	76.30.84.283.1		1
ELBO	ELBOW LR90 BW SCH 30		ASME B16.9	ASTM A420 GR. WPL6	8	76.30.39.283.1		4
FLAN	FLANGE 150# RFWN SCH 30		ASME B16.5	ASTM A350-LF2	8	76.70.82.058.1		4
FLAN	FLANGE 150# RFWN SCH 40		ASME B16.5	ASTM A350-LF2	4	76.70.82.114.1		2
FLAN	FLANGE 600# RFWN SCH 40		ASME B16.5	ASTM A105	2	76.62.86.076.1		1
REDU	REDU. ECC. BW SCH 40 X SCH 30		ASME B16.9	ASTM A420 GR. WPL6	8X4	76.31.70.324.1		2
TEE	TEE BW SCH30		ASME B16.6	ASTM A420 GR. WPL6	8X8	76.30.84.283.1		1
TUBI	PIPE SMLS SCH 30			ASTM A333 GR.6	8	74.13.17.093.1	11.5m	
TUBI	PIPE SMLS SCH 40			ASTM A333 GR.6	2	74.13.17.069.1	0.6m	
VALV	GATE VALVE 150# RF	WITH EXTENDED HANDWHEEL OPERATOR FLANGED,ASTM A352, A350 BODY WITH AISI 316 (L) TRIM			8	70.00.00.002.1		2

Appendix 4: Piping and Equipment Audit Case Study Data *(Continued)*

8-P14547-61170X

Type	Detail	Comment	Standard	Material	N.S. (INS)	Item Code	Length	Quantity
ELBO	ELBOW LR90 BW SCH 80		ASME B16.9	ASTM A420 GR. WPL-6	8	76.31.39.288.1		1
FBLI	FLANGE BLIND 600# RF		ASME B16.5	ASTM A350 GR.LF2	1	76.62.12.566.1		1
FLAN	FLANGE 600# RFWN SCH 80		ASME B16.5	ASTM A350 GR.LF2	8	76.62.86.168.1		3
INST	FLANGE CONNECTOR, 600#				0.5	70.00.00.007.1		1
INST	MONOFLANGE				1.5	70.00.00.006.1		1
INST	PRESSURE INDICATOR				0.5	70.00.00.008.1		1
INST	TEMPERATURE INDICATOR				0.5	70.00.00.009.1		1
TUBI	PIPE SMLS SCH 80			ASTM A333 GR.6	8	74.13.17.098.1	1.4m	
VALV	GLOBE VALVE 600# RF WITH EXTENDED STEM	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			1	70.00.00.005.1		1
VALV	CHECK VALVE SWING 600# RF	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			8	77.10.30.115.1		1
VALV	FLANGED 600# RF GATE VALVE WITH EXTENDED STEM	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			8	70.00.00.004.1		1
VALV	GLOBE VALVE WITH EXTENDED GEAR OPERATOR 600# RF	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			8	70.00.00.003.1		1

Appendix 4: Piping and Equipment Audit Case Study Data *(Continued)*

30-P12019-61170X

Type	Detail	Comment	Standard	Material	N.S. (INS)	Item Code	Length	Quantity
ELBO	ELBOW 45 DEG. BW 19.05MM WT		ASME B16.9	ASTM A420 GR. WPL-6	30	76.31.38.527.1		2
ELBO	ELBOW LR90 BW 15.88MM WT		ASME B16.9	ASTM A420 GR. WPL-6	24	76.31.39.407.1		2
ELBO	ELBOW LR90 BW 19.05MM WT		ASME B16.9	ASTM A420 GR. WPL-6	30	76.31.39.527.1		8
ELBO	ELBOW LR90 BW SCH 40		ASME B16.9	ASTM A420 GR. WPL-6	2	76.31.39.208.1		5
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A350 GR.LF2	24X0.5	76.80.37.702.1		1
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A350 GR.LF2	2X1	70.00.00.010.1		1
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A350 GR.LF2	30X1.5	70.00.00.011.1		3
INST	MONOFLANGE SINGLE BLOCK BALL 600# RF				1.5	70.00.00.013.1		2
INST	MONOFLANGE SINGLE BLOCK BALL 600# RF				0.5	70.00.00.014.1		1
INST	TEMPERATURE INDICATOR				0.5	70.00.00.009.1		1
FBLI	FLANGE BLIND 600# RF		ASME B16.5	ASTM A350 GR.LF2	1	76.62.12.566.1		1
FLAN	FLANGE 600# RF WN 19.05MM WT		ASME B16.47 SERIES A	ASTM A350 GR.LF2	30	76.62.83.220.1		4
FLAN	FLANGE 600# RFWN 15.88 MM WT		ASME B16.5	ASTM A350 GR.LF2	24	76.62.86.030.1		1
FLAN	FLANGE 600# RFWN SCH 40		ASME B16.5	ASTM A350 GR.LF2	2	76.62.86.076.1		2
PCOM	COVER FLANGE FOR FLANGED THERMOWELLS 600# RF			AISI 316	1.5X0.5	70.00.00.016.1		1
PCOM	THERMOWELL STRAIGHT, FLANGED 150#-900# RF 230MM LONG 316 SS			AISI 316	1.5	70.00.00.017.1		1
PCOM	SPACER RING TYPE 600# RF			ASTM A516 GR.60	30	70.00.00.015.1		1
PCOM	SPECTACLE BLIND FLANGE 600# RF			ASTM A516 GR.60	2	76.88.34.420.1		1
TUBI	PIPE SMLS SCH 40			ASTM A333 GR.6	2	74.13.17.059.1	7.5m	
TUBI	PIPE WELDED 19.05MM WT			ASTM A671 GR.CC65	30	74.30.09.377.1	38m	
TUBI	PIPE WELDED 15.88MM WT			ASTM A671 GR.CC65	24	74.30.09.282.1	3.6m	
VALV	BUTTERFLY VALVE TRIPLE ECC TYPE 600# RF	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			30	77.57.54.126.9		1
VALV	FLANGED 600# RF GATE VALVE WITH EXTENDED STEM	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			24	70.00.00.012.1		1
VALV	GLOBE VALVE 600# RF	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			2	77.30.30.108.1		1
VALV	VALVE GATE FLANGED 600# RF	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			2	77.30.20.108.1		1

Appendix 4: Piping and Equipment Audit Case Study Data *(Continued)*

30-P13001-61170X

Type	Detail	Comment	Standard	Material	N.S. (INS)	Item Code	Length	Quantity
ELBO	ELBOW LR90 BW 19.05MM WT		ASME B16.9	ASTM A420 GR. WPL-6	30	76.31.39.527.1		3
ELBO	ELBOW LR90 BW SCH 40		ASME B16.9	ASTM A420 GR. WPL-6	2	76.31.39.208.1		2
ELBO	ELBOW LR90 BW SCH 40		ASME B16.9	ASTM A420 GR. WPL-6	4	76.31.39.246.1		2
FBLI	FLANGE BLIND 600# RF		ASME B16.5	ASTM A350 GR.LF2	2	76.62.12.570.1		1
FBLI	FLANGE BLIND 600# RF		ASME B16.5	ASTM A350 GR.LF2	1	76.62.12.566.1		3
FLAN	FLANGE 600# RF WN 19.05MM WT		ASME B16.47 SERIES A	ASTM A350 GR.LF2	30	76.62.83.220.1		2
FLAN	FLANGE 600# RFWN SCH 40		ASME B16.5	ASTM A350 GR.LF2	4	76.62.86.080.1		4
FLAN	FLANGE 600# RFWN SCH 40		ASME B16.5	ASTM A350 GR.LF2	3	76.62.86.078.1		2
FLAN	FLANGE 600# RFWN SCH 40		ASME B16.5	ASTM A350 GR.LF2	2	76.62.86.076.1		7
FLAN	FLANGE 600# RFWN SCH 80		ASME B16.5	ASTM A350 GR.LF2	1	76.62.86.156.1		1
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A350 GR.LF2	30X1.5	70.00.00.011.1		1
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A350 GR.LF2	30X0.5	70.00.00.030.1		1
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A350 GR.LF2	2X0.5	76.80.37.701.1		1
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A350 GR.LF2	4X1	76.80.37.707.1		2
OLET	WELDOLET SCH 40 X SCH 40			ASTM A350 GR.LF2	4X2	70.00.00.026.1		1
INST	MONOFLANGE				0.5	70.00.00.024.1		1
INST	MONOFLANGE SINGLE BLOCK BALL 600# RF			ASTM A350 GR.LF3	0.5	70.00.00.014.1		1
INST	PRESSURE SAFETY VALVE 600#RF x 150#RF				4X3	70.00.00.027.1		2
INST	PRESSURE INDICATOR			AISI 316	0.5	70.00.00.008.1		1
INST	THERMOCOUPLE			AISI 316	0.5	70.00.00.033.1		1
PCOM	COVER FLANGE FOR FLANGED THERMOWELLS 600# RF			AISI 316	1.5X0.5	70.00.00.016.1		1
PCOM	SPECTACLE BLIND FLANGE 600# RF			ASTM A516 GR.60	2	76.88.34.420.1		1
REDU	REDU. CONC. BW SCH 40		ASME B16.9	ASTM A420 GR. WPL-6	4X3	76.31.70.275.1		2
REDU	REDU. CONC. BW SCH 40 x SCH 80		ASME B16.9	ASTM A420 GR. WPL-6	2X1	76.31.70.198.1		1
TEE	TEE BW SCH 40		ASME B16.9	ASTM A420 GR. WPL-6	4X4	76.31.84.246.1		1
TEE	TEE BW SCH 40		ASME B16.9	ASTM A420 GR. WPL-6	2X2	76.31.84.208.1		2
TUBI	PIPE SMLS SCH 40			ASTM A333 GR.6	4	74.13.17.075.1	5.9m	
TUBI	PIPE SMLS SCH 40			ASTM A333 GR.6	2	74.13.17.059.1	1.4m	
TUBI	PIPE WELDED 19.05MM WT			ASTM A671 GR.CC65	30	74.13.17.377.1	8.2m	
VALV	GLOBE VALVE 600# RF WITH EXTENDED STEM	ASTM A351 -CF8M BODY WITH AISI 316 TRIM			2	70.00.00.020.1		2
VALV	CHECK VALVE SWING 600# RF	ASTM A352 , A350 BODY WITH AISI 316 (L) TRIM			2	70.00.00.022.1		1
VALV	FLANGED 600# RF GATE VALVE WITH EXTENDED STEM	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			4	70.00.00.028.1		2
VALV	FLANGED 600# RF GATE VALVE WITH EXTENDED STEM	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			2	70.00.00.021.1		2
VALV	FLANGED 600# RF GATE VALVE WITH EXTENDED STEM	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			1	70.00.00.023.1		3
VALV	GEAR OPERATED BUTTERFLY VALVE 600# RF	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			30	70.00.00.019.1		1

Appendix 4: Piping and Equipment Audit Case Study Data *(Continued)*

30-P13002-61170X

Type	Detail	Comment	Standard	Material	N.S. (INS)	Item Code	Length	Quantity
ELBO	ELBOW 45 DEG. BW SCH 40		ASME B16.9	ASTM A420 GR. WPL-6	2	76.31.38.208.1		1
ELBO	ELBOW LR90 BW 19.05MM WT		ASME B16.9	ASTM A420 GR. WPL-6	30	76.31.39.527.1		8
ELBO	ELBOW LR90 BW SCH 40		ASME B16.9	ASTM A420 GR. WPL-6	2	76.31.39.208.1		22
ELBO	ELBOW LR90 BW SCH 80		ASME B16.9	ASTM A420 GR. WPL-6	1	76.31.39.180.1		2
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A350 GR.LF2	30X1.5	70.00.00.011.1		1
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A350 GR.LF2	30X0.5	70.00.00.030.1		4
OLET	BRANCH FITTING FLANGED-OUTLET 150# RF SCH 40 X SCH 80			ASTM A333 GR.6	2X1	70.00.00.040.1		1
FBLI	FLANGE BLIND 600# RF		ASME B16.5	ASTM A350 GR.LF2	1	76.62.12.566.1		3
FBLI	FLANGE BLIND 600# RF		ASME B16.5	ASTM A350 GR.LF2	2	76.62.12.570.1		2
FLAN	FLANGE 600# RF WN 19.05MM WT		ASME B16.47 SERIES A	ASTM A350 GR.LF2	30	76.62.83.220.1		7
FLAN	FLANGE 600# RFWN SCH 40		ASME B16.5	ASTM A350 GR.LF2	2	76.62.86.076.1		20
FLAN	FLANGE 600# RFWN SCH 80		ASME B16.5	ASTM A350 GR.LF2	1	76.62.86.156.1		4
INST	MONOFLANGE				0.5	70.00.00.024.1		2
INST	MONOFLANGE SINGLE BLOCK BALL 600# RF				0.5	70.00.00.014.1		2
INST	PRESSURE INDICATOR			AISI 316	0.5	70.00.00.008.1		2
INST	PRESSURE SAFETY VALVE				2X1	70.00.00.048.1		2
INST	THERMOCOUPLE			AISI 316	0.5	70.00.00.033.1		1
PCOM	COVER FLANGE FOR FLANGED THERMOWELLS 600# RF			AISI 316	1.5X0.5	70.00.00.016.1		1
PCOM	SPACER RING TYPE 600# RF			ASTM A516 GR.60	30	70.00.00.015.1		3
PCOM	SPECTACLE BLIND FLANGE 600# RF			ASTM A516 GR.60	2	76.88.34.420.1		2
REDU	REDU. CONC. BW SCH 40 x SCH 80		ASME B16.9	ASTM A420 GR. WPL-6	2X1	76.31.70.198.1		4
TEE	TEE BW 19.05MM WT		ASME B16.9	ASTM A420 GR. WPL-6	30X30	76.31.84.527.1		1
TEE	TEE BW SCH 40		ASME B16.9	ASTM A420 GR. WPL-6	2X2	76.31.84.208.1		7
TUBI	PIPE SMLS SCH 40			ASTM A333 GR.6	2	74.13.17.059.1	43.9m	
TUBI	PIPE SMLS SCH 80			ASTM A333 GR.6	1	74.13.17.037.1	0.2m	
TUBI	PIPE WELDED 19.05MM WT			ASTM A671 GR.CC65	30	74.30.09.377.1	24.6m	
VALV	GLOBE VALVE 600# RF WITH EXTENDED STEM	ASTM A351 -CF8M BODY WITH AISI 316 TRIM			2	70.00.00.020.1		4
VALV	CHECK VALVE SWING 600# RF	ASTM A352 , A350 BODY WITH AISI 316 (L) TRIM			2	70.00.00.022.1		2
VALV	FLANGED 600# RF GATE VALVE WITH EXTENDED STEM	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			2	70.00.00.021.1		6
VALV	FLANGED 600# RF GATE VALVE WITH EXTENDED STEM	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			1	70.00.00.023.1		2
VALV	GEAR OPERATED BUTTERFLY VALVE 600# RF	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			30	70.00.00.019.1		2

Appendix 4: Piping and Equipment Audit Case Study Data *(Continued)*

30-P13003-61170X

Type	Detail	Comment	Standard	Material	N.S. (INS)	Item Code	Length	Quantity
ELBO	ELBOW LR90 BW 15.88MM WT		ASME B16.9	ASTM A420 GR. WPL-6	24	76.31.39.407.1		5
ELBO	ELBOW LR90 BW 19.05MM WT		ASME B16.9	ASTM A420 GR. WPL-6	30	76.31.39.527.1		3
ELBO	ELBOW LR90 BW SCH 40		ASME B16.9	ASTM A420 GR. WPL-6	2	76.31.39.208.1		9
ELBO	ELBOW LR90 BW SCH 40		ASME B16.9	ASTM A420 GR. WPL-6	4	76.31.39.246.1		9
ELBO	ELBOW LR90 BW SCH 80		ASME B16.9	ASTM A420 GR. WPL-6	1.5	76.31.39.188.1		2
ELBO	ELBOW LR90 BW SCH 80		ASME B16.9	ASTM A420 GR. WPL-6	0.5	76.31.39.168.1		1
FBLI	FLANGE BLIND 600# RF		ASME B16.5	ASTM A350 GR.LF2	1.5	76.62.12.568.1		1
FBLI	FLANGE BLIND 600# RF		ASME B16.5	ASTM A350 GR.LF2	1	76.62.12.566.1		3
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A350 GR.LF2	24X1.5	76.80.37.714.1		2
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A350 GR.LF2	24X0.5	76.80.37.702.1		1
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A350 GR.LF2	4X1	76.80.37.707.1		2
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A333 GR.6	2X1	70.00.00.010.1		1
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A350 GR.LF2	30X0.5	70.00.00.030.1		1
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A350 GR.LF2	1.5X0.5	70.00.00.055.1		2
OLET	BRANCH FITTING FLANGED-OUTLET 600# RF SCH 80			ASTM A350 GR.LF2	2X0.5	76.80.37.701.1		2
FLAN	FLANGE 600# RF WN 19.05MM WT		ASME B16.47 SERIES A	ASTM A350 GR.LF2	30	76.62.83.220.1		6
FLAN	FLANGE 600# RFWN 15.88 MM WT		ASME B16.5	ASTM A350 GR.LF2	24	76.62.86.030.1		3
FLAN	FLANGE 600# RFWN SCH 40		ASME B16.5	ASTM A350 GR.LF2	2	76.62.86.076.1		7
FLAN	FLANGE 600# RFWN SCH 40		ASME B16.5	ASTM A350 GR.LF2	4	76.62.86.080.1		6
FLAN	FLANGE 600# RFWN SCH 80		ASME B16.5	ASTM A350 GR.LF2	1.5	76.62.86.158.1		4
FLAN	FLANGE 600# RFWN SCH 80		ASME B16.5	ASTM A350 GR.LF2	0.5	76.62.86.152.1		2
INST	MONOFLANGE				0.5	70.00.00.024.1		2
INST	MONOFLANGE SINGLE BLOCK BALL 600# RF				0.5	70.00.00.014.1		2
INST	PRESSURE INDICATOR			AISI 316	0.5	70.00.00.008.1		2
INST	THERMOCOUPLE			AISI 316	0.5	70.00.00.033.1		1
INST	VALVE CONTROL 600# RF				24	70.00.00.063.1		1
PCOM	COVER FLANGE FOR FLANGED THERMOWELLS 600# RF			AISI 316	1.5X0.5	70.00.00.016.1		1
PCOM	SPACER RING TYPE 600# RF			ASTM A516 GR.60	24	76.88.17.194.1		1
PCOM	SPACER RING TYPE 600# RF			ASTM A516 GR.60	30	70.00.00.015.1		2
PCOM	SPACER RING TYPE 600# RF			ASTM A516 GR.60	4	78.88.17.174.1		1
PCOM	SPECTACLE BLIND FLANGE 600# RF			ASTM A516 GR.60	2	76.88.34.420.1		1
PCOM	SPECTACLE BLIND FLANGE 600# RF			ASTM A516 GR.60	4	76.88.34.424.1		1
REDU	REDU. CONC. BW 19.05MM WT x 15.88MM WT		ASME B16.9	ASTM A420 GR. WPL-6	30X24	76.31.60.740.1		1
REDU	REDU. CONC. BW SCH 40 x SCH 80		ASME B16.9	ASTM A420 GR. WPL-6	2X1.5	76.31.70.206.1		2
TEE	TEE BW 19.05MM WT		ASME B16.9	ASTM A420 GR. WPL-6	30X30	76.31.84.527.1		1
TUBI	PIPE SMLS SCH 40			ASTM A333 GR.6	2	74.13.17.059.1	11.9m	
TUBI	PIPE SMLS SCH 40			ASTM A333 GR.6	4	74.13.17.075.1	12.7m	
TUBI	PIPE SMLS SCH 80			ASTM A333 GR.6	1.5	74.13.17.053.1	0.5m	
TUBI	PIPE SMLS SCH 80			ASTM A333 GR.6	0.5	74.13.17.021.2	0.1m	
TUBI	PIPE WELDED 19.05MM WT			ASTM A671 GR.CC65	30	74.30.09.377.1	9.3m	
TUBI	PIPE WELDED 15.88MM WT			ASTM A671 GR.CC65	24	74.30.09.282.1	10.7m	
VALV	GLOBE VALVE 600# RF WITH EXTENDED STEM	ASTM A351 -CF8M BODY WITH AISI 316 TRIM			2	70.00.00.020.1		1
VALV	CHECK VALVE PISTON 600# RF	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			1.5	77.10.30.107.1		1
VALV	FLANGED 600# RF GATE VALVE WITH EXTENDED STEM	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			2	70.00.00.021.1		3
VALV	FLANGED 600# RF GATE VALVE WITH EXTENDED STEM	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			4	70.00.00.028.1		2
VALV	FLANGED GLOBE VALVE WITH EXTENDED HANDWHEEL 600# RF	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			1.5	70.00.00.059.1		1
VALV	GEAR OPERATED BUTTERFLY VALVE 600# RF	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			30	70.00.00.019.1		2
VALV	GLOBE VALVE WITH EXTENDED GEAR OPERATOR 600# RF	ASTM A350 GR. LF2 BODY WITH AISI 316 (L) TRIM			4	70.00.00.051.1		2
VALV	VALVE NEEDLE STRAIGHT NPT				0.5	70.00.00.060.1		1

Appendix 5: Sample Precision and Recall Calculations

This example explains the basic concept of precision and recall evaluation using a classifier in the piping and equipment audit case study. The classifier in the case study has scores for its predictions, and the predictions have labels (Positive and Negative). To carry out a model-wide evaluation, the performance of the classifier must be examined at different classification thresholds. The following are the 5 steps required for this type of calculation.

STEP 1: Sort the data by score

Using classifier 1 observed scores as shown in Figure 6.2, the data is sorted by scores as in Table A.3: Model-wide evaluation for classifier 1 on task 1

STEP 2: Decide the threshold values

Considering the observed scores range between 0 and 100, threshold values between these intervals are used. Also, a constant interval of 10 is chosen to make the calculations straight forward.

STEP 3: Determine predicted label for each threshold value

Using the defined threshold values, predicted labels for the different thresholds are generated using the logic that threshold values less than or equal to an observed score will be positively predicted. Similarly, threshold values more than the observed score will be negatively predicted. This is shown in Table A.3

STEP 4: Generate confusion matrix

The elements of a confusion matrix as shown in Figure 4.8(A) include false negative (FN), true positive (TP), false positive (FP) and true negative (TN). Their dot products as shown in Figure 4.8(B) also suffice. Using this concept, the products of the observed and predicted labels are counted and summarised as 'Sums' in Table A.3

Appendix 5: Sample Precision and Recall Calculations *(Continued)*

STEP 5: Compute evaluation scores.

Using Equations 4.1 to 4.5 described in the Section 4.3(j), evaluation measures for precision and recall are calculated as shown in Table A.3. Table A.4 shows identical computation for classifier 1 on task 2 in Figure 6.5.

Appendix 5: Sample Precision and Recall Calculations *(Continued)*

Table A.3: Model-wide evaluation for classifier 1 on task 1

Classifier 1			Predicted labels at different thresholds										
Observed	Score		TH=100	TH=90	TH=80	TH=70	TH=60	TH=50	TH=40	TH=30	TH=20	TH=10	TH=0
P	19		N	N	N	N	N	N	N	N	N	P	P
P	93.75		N	P	P	P	P	P	P	P	P	P	P
P	93.75		N	P	P	P	P	P	P	P	P	P	P
P	93.75		N	P	P	P	P	P	P	P	P	P	P
P	93.75		N	P	P	P	P	P	P	P	P	P	P
P	99.75		N	P	P	P	P	P	P	P	P	P	P
P	100		P	P	P	P	P	P	P	P	P	P	P
N	100		P	P	P	P	P	P	P	P	P	P	P
		FN	6	1	1	1	1	1	1	1	1	0	0
	[Sums]	**TP**	1	6	6	6	6	6	6	6	6	7	7
		FP	1	1	1	1	1	1	1	1	1	1	1
		TN	0	0	0	0	0	0	0	0	0	0	0
		Recall	0.142857	0.857143	0.857143	0.857143	0.857143	0.857143	0.857143	0.857143	0.857143	1	1
		Precision	0.5	0.857143	0.857143	0.857143	0.857143	0.857143	0.857143	0.857143	0.857143	0.875	0.875
		F1	0.222222	0.857143	0.857143	0.857143	0.857143	0.857143	0.857143	0.857143	0.857143	0.933333	0.933333
		FPR	1	1	1	1	1	1	1	1	1	1	1

Appendix 5: Sample Precision and Recall Calculations *(Continued)*

Table A.4: Model-wide evaluation for classifier 1 on task 2

Classifier 1			Predicted labels at different thresholds										
Observed	Score		TH=100	TH=90	TH=80	TH=70	TH=60	TH=50	TH=40	TH=30	TH=20	TH=10	TH=0
P	19		N	N	N	N	N	N	N	N	N	P	P
P	93.75		N	P	P	P	P	P	P	P	P	P	P
P	93.75		N	P	P	P	P	P	P	P	P	P	P
N	100		P	P	P	P	P	P	P	P	P	P	P
N	100		P	P	P	P	P	P	P	P	P	P	P
N	100		P	P	P	P	P	P	P	P	P	P	P
N	100		P	P	P	P	P	P	P	P	P	P	P
N	100		P	P	P	P	P	P	P	P	P	P	P
		FN	3	1	1	1	1	1	1	1	1	0	0
	[Sums]	**TP**	0	2	2	2	2	2	2	2	2	3	3
		FP	5	5	5	5	5	5	5	5	5	5	5
		TN	0	0	0	0	0	0	0	0	0	0	0
		Recall	0	0.666667	0.666667	0.666667	0.666667	0.666667	0.666667	0.666667	0.666667	1	1
		Precision	0	0.285714	0.285714	0.285714	0.285714	0.285714	0.285714	0.285714	0.285714	0.375	0.375
		F1	$\propto$	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.545455	0.545455
		FPR	1	1	1	1	1	1	1	1	1	1	1

Appendix 6: Precision and Recall Precrec Commands

Precrec is a software package for computing Receiver Operating Characteristics (ROC) and Precision -Recall (PRC) Curves (Saito and Rehmsmmeier, 2017). It enables a robust performance evaluation of binary classifiers and returns result for a fairly large dataset within seconds. Precrec was used in this research for plotting ROC and precision-recall curves as well as calculating the areas under the curves. The command functions applied are described as follows:

Starting up

```
# Loading of  the required libraries

library(precrec)
library(ggplot2)
```

Classifier 1/Task1

```
# Entering the scores for classifier 1
scores1 <- c(93.75, 93.75, 19, 100, 100, 99.75, 93.75, 93.75)
# Entering the prediction values for the scores of classifier 1 (P=1, N=0)
labels1 <- c(1, 1, 1, 1, 0, 1, 1, 1)
# Reformat input data for performance evaluation calculation
msmdat <- mmdata(scores1, labels1)
# Calculating evaluation measures including ROC and Precision-Recall curves
mscurves <- evalmod(msmdat)
# Plot performance evaluation measures - ROC and Precision-Recall curves
autoplot(mscurves)
#Calculate area under the curves (aucs)
auc(mscurves)
```

Returned aucs for classifier1/Task1

modnames	dsids	curvetypes	aucs
m1	1	ROC	0.07142857
m1	1	PRC	0.73052924

Appendix 6: Precision and Recall Precrec Commands *(Continued)*

Classifier 2/Task 1

```
scores1 <- c(100, 100, 100, 100, 100, 100, 99.75, 99.75)
labels1 <- c(1, 1, 1, 1, 0, 1, 1, 1)
msmdat <- mmdata(scores1, labels1)
mscurves <- evalmod(msmdat)
autoplot(mscurves)
auc(mscurves)
```

Returned aucs for classifier2/Task1

modnames	dsids	curvetypes	aucs
m1	1	ROC	0.3571429
m1	1	PRC	0.8398549

Classifier 3/Task 1 (Classifier 3 = Classifier 4 = Classifier 5)

```
scores1 <- c(100, 100, 100, 100, 100, 100, 100, 100)
labels1 <- c(1, 1, 1, 1, 0, 1, 1, 1)
msmdat <- mmdata(scores1, labels1)
mscurves <- evalmod(msmdat)
autoplot(mscurves)
auc(mscurves)
```

Returned aucs for classifier3/Task1

modnames	dsids	curvetypes	aucs
m1	1	ROC	0.5
m1	1	PRC	0.875

Appendix 6: Precision and Recall Precrec Commands *(Continued)*

Classifier 1/Task 2

```
scores1 <- c(93.75, 93.75, 19, 100, 100, 100, 100, 100)
labels1 <- c(1, 1, 1, 0, 0, 0, 0, 0)
msmdat <- mmdata(scores1, labels1)
mscurves <- evalmod(msmdat)
autoplot(mscurves)
auc(mscurves)
```

Returned aucs for classifier1/Task2

modnames	dsids	curvetypes	aucs
m1	1	ROC	0.0000000
m1	1	PRC	0.2166606

Classifier 2/Task2 (Classifier 2 = Classifier 3 = Classifier 4 = Classifier 5)

```
scores1 <- c(100, 100, 100, 100, 100, 100, 100, 100)
labels1 <- c(1, 1, 1, 0, 0, 0, 0, 0)
msmdat <- mmdata(scores1, labels1)
mscurves <- evalmod(msmdat)
autoplot(mscurves)
auc(mscurves)
```

Returned aucs for classifier2/Task2

modnames	dsids	curvetypes	aucs
m1	1	ROC	0.500
m1	1	PRC	0.375

Appendix 6: Precision and Recall Precrec Commands *(Continued)*

Plotting F1 Score for classifiers on Task 1

```
s1 <- c(93.75, 93.75, 19, 100, 100, 99.75, 93.75, 93.75)
s2 <- c(100, 100, 100, 100, 100, 100, 99.75, 99.75)
s3 <- c(100, 100, 100, 100, 100, 100, 100, 100)
# Joining scores of multiple classifiers so that they can be plotted on same graph
scores1 <- join_scores(s1, s2, s3)
labels1 <- c(1, 1, 1, 1, 0, 1, 1, 1)
# Specifying classifier names and corresponding dataset ID
mmmdat <- mmdata(scores1, labels1, modnames = c("classifier1", "classifier2", "classifier3"), dsids
= c(1, 2, 3))
# Calculating the basic evaluation measures instead of performing ROC and Precision-Recall
calculations
mmpoins <- evalmod(mmmdat, mode = "basic")
#Plotting normalized ranks vs. Matthews correlation coefficient and F-score
autoplot(mmpoins, c("mcc", "fscore"))
```

Plotting F1 Score for classifiers on Task 2

```
s1 <- c(93.75, 93.75, 19, 0, 100, 100, 100, 100)
s2 <- c(100, 100, 100, 100, 100, 100, 100, 100)
scores1 <- join_scores(s1, s2)
labels1 <- c(1, 1, 1, 0, 0, 0, 0, 0)
mmmdat <- mmdata(scores1, labels1, modnames = c("classifier1", "classifier2"), dsids = c(1, 2))
mmpoins <- evalmod(mmmdat, mode = "basic")
autoplot(mmpoins, c("mcc", "fscore"))
```

Appendix 6: Precision and Recall Precrec Commands *(Continued)*

Plotting F1 Score for semantic classifier on Task 1

```
# A dataset with 7 positives and 1 negative for the perfect performance level
samps1 <- create_sim_samples(1, 7, 1, "perf"))
msmdat3 <- mmdata(samps1[["scores"]], samps1[["labels"]], modnames = samps1[["modnames"]])
mscurves <- evalmod(msmdat3)
autoplot(mscurves)
mmpoins <- evalmod(msmdat3, mode = "basic")
autoplot(mmpoins, c("mcc", "fscore"))
```

Plotting F1 Score for semantic classifier on Task 2

```
A dataset with 3 positives and 5 negatives for the perfect performance level
samps1 <- create_sim_samples(1, 3, 5, "perf"))
msmdat3 <- mmdata(samps1[["scores"]], samps1[["labels"]], modnames = samps1[["modnames"]])
mscurves <- evalmod(msmdat3)
autoplot(mscurves)
mmpoins <- evalmod(msmdat3, mode = "basic")
autoplot(mmpoins, c("mcc", "fscore"))
```

Appendix 7: Data Mapping Using dot15926 Editor

This section provides the steps for mapping to RDF the steel section and 3D point cloud summary data in Section 6.2.1 using dot15926 Editor. dot15926 Editor is an open source software developed by TechInvestLab.ru (2013). This guide is valid for version 1.5beta of the software. The data sets required for the replication of the described data transformations are provided in Akinyemi (2018a).The data mapping process involves preprocessing data in spreadsheets and using dot15926 Editor's built-in spreadsheet import function for data transformation. The important steps for the process are detailed as follows:

STEP 1: Set up of data project in dot15926 Editor

First, dot15926 Editor is downloaded from http://techinvestlab.ru/dot15926Editor. In setting up the data project, consideration is given to the requirement that entities in the sources need to be related to corresponding ISO 15926 entities. As a result, the following ISO 15926 resources are added to the data project.

- ISO 15926 Part 2 Type Entities (accessible from dot15926 Editor's menu)
- ISO 15926 Part 4 RDL (downloaded from http://rds.posccaesar.org/downloads/PCA-RDL.owl.zip and registered in Editor settings for accessibility from its menu)
- Template Information Patterns based on ISO 15926 Parts 7 and 8 (tips25062014.patt) - available with the software and can be added to a new project by copying to *<installation_folder>/patterns* folder
- Templates from ISO 15926 Information Patterns (IIP) project that lead to the development of the latter (iip_fullset_20140131_PCA_dm.owl) - available with the software and can be added to a new project by adding it as a new data file

For the case study in this research, a new project is created in dot15926 Editor and named 'Steel_Waste_Management'. Thereafter, the following data files are added as shown in Figure A.4.

- information patterns file for the project (Steel_Waste_Patterns.patt)
- template information patterns (tips25062014.patt and iip_fullset_20140131_PCA_dm.owl)
- the registered ISO 15926 Part 4 RDL (PCA RDL)
- a new data file for the actual project data (Steel_Data)
- a new data file for project specific reference data library (Steel_RDL)

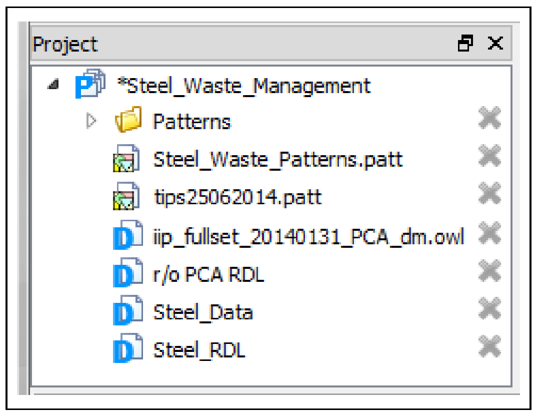


Figure A.4: Project panel of dot15926 Editor

In the project properties shown in Figure A.5, annotation property 'hasLocalId' is registered for the project, and the annotation properties in the PCA RDL are copied to the new project's properties. This set is used as a baseline for project annotation properties. Additional annotations e.g. 'hasImage' are added to the list as required. Similarly, additional object properties e.g. 'hasLocation' are added to the Editor's default object properties, 'subClassOf' and 'type', as required.

Appendix 7: Data Mapping Using dot15926 Editor *(Continued)*

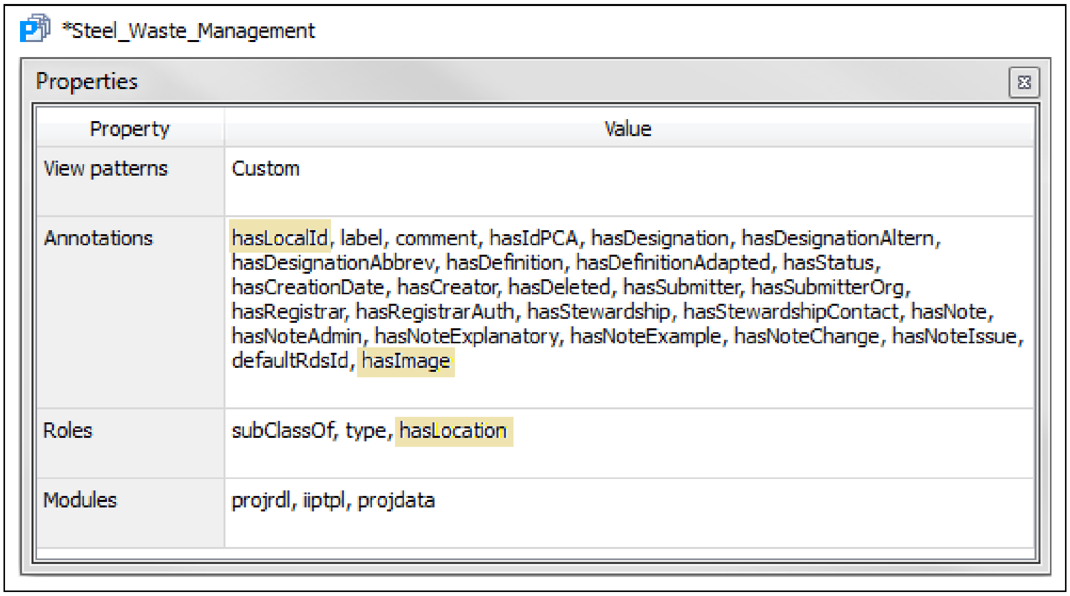


Figure A.5: Project annotation and roles view in dot15926 Editor

Another setup requirement is that the modules for the data project are set as shown in Figure A.6. Project reference data library (projrdl) module is connected to the data file for project specific reference data library (Steel_RDL); ISO 15926 information patterns template (iiptpl) module is connected to the template information patterns iip_fullset_20140131_PCA_dm.owl; and project data (projdata) module connected to project data file, Steel_Data.

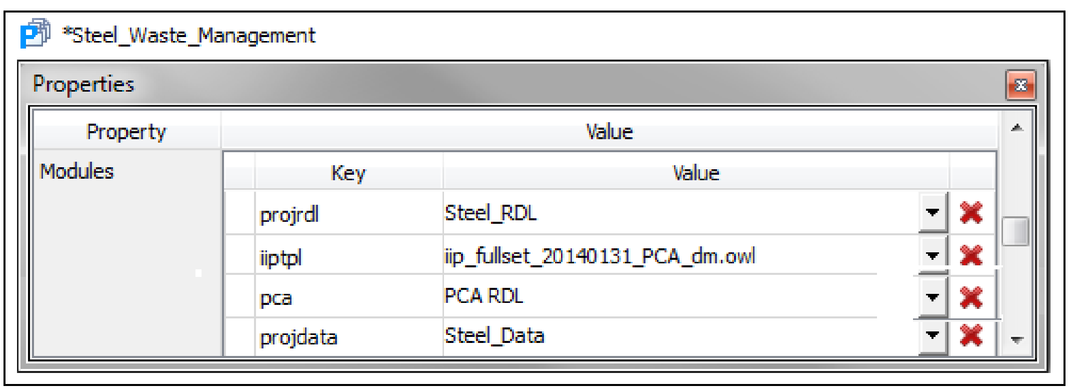


Figure A.6: Project modules view in dot15926 Editor

Lastly, the two namespaces that will be used for mapping, <http://steel.waste.com/rdl/> (for project specific reference data) and <http://steel.waste.com/project/> (for the project data), are generated and registered as properties of the data files that will receive the transformed data. Figure A.7 represents the registration for this case study.

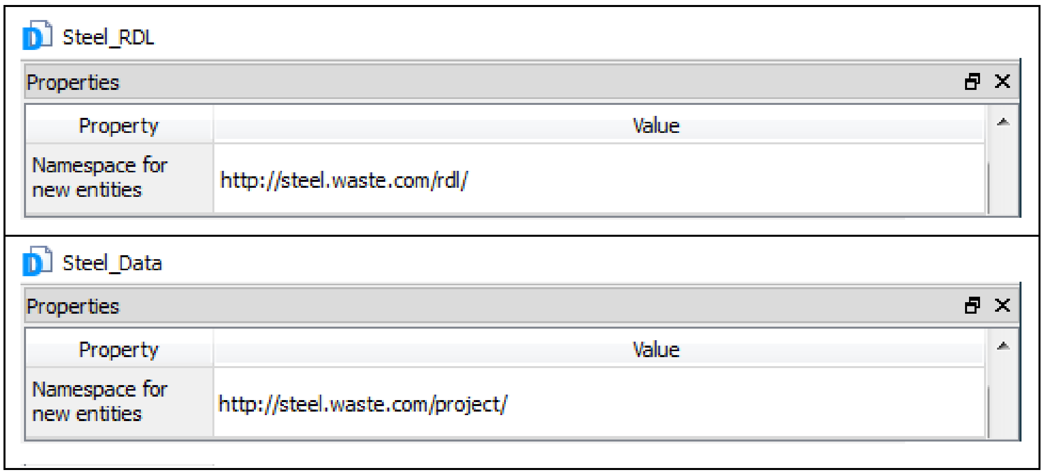


Figure A.7: Namespace registration for new data files in dot15926 Editor

STEP 2: Preparation of data sources

First, a project specific reference data library is required for the entities in the sources. In this case study, there are eleven unique entities in the sources. These are listed in Column D of Figure A.8 by their specifications. To prepare the project reference data, the grey coloured columns in Figure A.8 are introduced. Cell A2 in Figure A.8 contains the namespace that is used in forming URIs for the new reference data entities. The URIs are formed by concatenating the namespace in Cell A2 with the local IDs in Column E of Figure A.8. The Local IDs in Column E are derived from the specifications registered in Column D of Figure A.8 using the excel formula:

=SUBSTITUTE(SUBSTITUTE(D2, "x", "-"), " ","") (A7.1)

Description and specification in Figure A.8 are preserved to import them as annotation properties. Column F in Figure A.8 contains the Part 2 type of the reference data entities, Column G contains the URIs of the PCA RDL superclasses, and Column H their names for easy referencing. Columns I and J contain metadata for the project RDL entities.

Appendix 7: Data Mapping Using dot15926 Editor *(Continued)*

	A	B	C	D	E
1	RDL namespace	URI	Description	Specification	Local ID
2	http://steel.waste.com/rdl/	http://steel.waste.com/rdl/38-5SG	Grating	38x5 SG	38-5SG
3		http://steel.waste.com/rdl/CHS168.3-5.6	Circular Hollow Sectior	CHS 168.3x5.6	CHS168.3-5.6
4		http://steel.waste.com/rdl/CHS219.1-5.0	Circular Hollow Sectior	CHS 219.1x5.0	CHS219.1-5.0
5		http://steel.waste.com/rdl/CHS323.9-5.6	Circular Hollow Sectior	CHS 323.9x5.6	CHS323.9-5.6
6		http://steel.waste.com/rdl/CHS355.69-8.0	Circular Hollow Sectior	CHS 355.69x8.0	CHS355.69-8.0
7		http://steel.waste.com/rdl/PFC260-75-28	Parrallel Flange Chann	PFC 260x75x28	PFC260-75-28
8		http://steel.waste.com/rdl/UB762-267-197	Universal Beam	UB 762x267x197	UB762-267-197
9		http://steel.waste.com/rdl/BD		Boat Deck	BD
10		http://steel.waste.com/rdl/CD		Cellar Deck	CD
11		http://steel.waste.com/rdl/MD		Main Deck	MD
12		http://steel.waste.com/rdl/3PC		3D Point Cloud	3PC

	F
1	Part 2 Type
2	http://rds.posccaesar.org/2008/02/OWL/ISO-15926-2_2003#ClassOfInanimatePhysicalObject
3	http://rds.posccaesar.org/2008/02/OWL/ISO-15926-2_2003#ClassOfInanimatePhysicalObject
4	http://rds.posccaesar.org/2008/02/OWL/ISO-15926-2_2003#ClassOfInanimatePhysicalObject
5	http://rds.posccaesar.org/2008/02/OWL/ISO-15926-2_2003#ClassOfInanimatePhysicalObject
6	http://rds.posccaesar.org/2008/02/OWL/ISO-15926-2_2003#ClassOfInanimatePhysicalObject
7	http://rds.posccaesar.org/2008/02/OWL/ISO-15926-2_2003#ClassOfInanimatePhysicalObject
8	http://rds.posccaesar.org/2008/02/OWL/ISO-15926-2_2003#ClassOfInanimatePhysicalObject
9	http://rds.posccaesar.org/2008/02/OWL/ISO-15926-2_2003#ClassOfInanimatePhysicalObject
10	http://rds.posccaesar.org/2008/02/OWL/ISO-15926-2_2003#ClassOfInanimatePhysicalObject
11	http://rds.posccaesar.org/2008/02/OWL/ISO-15926-2_2003#ClassOfInanimatePhysicalObject
12	http://rds.posccaesar.org/2008/02/OWL/ISO-15926-2_2003#RepresentationForm

	G	H	I	J
1	PCA RDL Superclass URI	PCA RDL Superclass	Date	Creator
2	http://posccaesar.org/rdl/RDS644309	GRATING PLATE	10/5/2017	A.G.
3	http://posccaesar.org/rdl/RDS992609	CHP 168.3 X 5.00 EN 10210-2	10/5/2017	A.G.
4	http://posccaesar.org/rdl/RDS980324	CHP 219.1 X 5.00 EN 10219-2	10/5/2017	A.G.
5	http://posccaesar.org/rdl/RDS981404	CHP 323.9 X 8.00 EN 10219-2	10/5/2017	A.G.
6	http://posccaesar.org/rdl/RDS981719	CHP 355.6 X 8.00 EN 10219-2	10/5/2017	A.G.
7	http://posccaesar.org/rdl/RDS1019294	PARALLEL FLANGE CHANNEL PROFILE	10/5/2017	A.G.
8	http://posccaesar.org/rdl/RDS1031894	UB 762 X 267 X 197	10/5/2017	A.G.
9	http://posccaesar.org/rdl/RDS288359	BOTTOM DECK	10/5/2017	A.G.
10	http://posccaesar.org/rdl/RDS288314	CELLAR DECK	10/5/2017	A.G.
11	http://posccaesar.org/rdl/RDS288404	MAIN DECK	10/5/2017	A.G.
12	http://posccaesar.org/rdl/RDS11368725	RASTER IMAGE FORMAT	10/5/2017	A.G.

Figure A.8: Preprocessed spreadsheet for case study reference data

To prepare the structural report for import, the grey coloured columns in Figure A.9 are introduced for preprocessing. Cells A2 and A5 in Figure A.9 contain the namespaces that are used in forming URIs for the data entities. The URIs for structural objects are formed by concatenating the namespace in Cell A2 and the unique IDs of structural objects in Column B (Figure A.9). Local IDs in Figure A.9's Columns F and M (classifiers) represent the parent objects of the structural objects. The local IDs in Column F are derived using the excel formula:

=SUBSTITUTE(SUBSTITUTE(G2, "x", "-"), " ","") (A7.2)

The local IDs in column M are derived by extracting the upper case letters in Column L using the excel formula:

=CONCAT(LEFT(L2,1),MID(L2,FIND(" ", L2,1)+1,1)) (A7.3)

To import the height and area attributes of the structural entities in Figure A.9, their units of measurement need to be separated. This is done using the following excel formulas:

cutlength in Column I =IF(ISBLANK(H2), "",LEFT(H2,2)) (A7.4)

area in Column K =IF(ISBLANK(J2), "",LEFT(J2,3)) (A7.5)

The URIs of the parent objects are formed by concatenating the namespace in Cell A5 (Figure A.9) and the corresponding values of the parent objects in Columns F and M (Figure A.9). Lastly, ID, Description and specification (Spec) of structural objects in Figure A.9 are preserved to import them as annotation properties, and Columns O and P in Figure A.9 contain metadata for the project entities.

Appendix 7: Data Mapping Using dot15926 Editor *(Continued)*

	A	B	C	D
1	Namespaces	ID	URI	Description
2	http://steel.waste.com/rdl/	B.01	http://steel.waste.com/project/B.01	Circular Hollow Section
3		B.02	http://steel.waste.com/project/B.02	Circular Hollow Section
4		B.03	http://steel.waste.com/project/B.03	Circular Hollow Section
5	http://steel.waste.com/project/	B.04	http://steel.waste.com/project/B.04	Circular Hollow Section
6		B.05	http://steel.waste.com/project/B.05	Circular Hollow Section
7		B.06	http://steel.waste.com/project/B.06	Circular Hollow Section
8		B.07	http://steel.waste.com/project/B.07	Circular Hollow Section
9		S.03	http://steel.waste.com/project/S.03	Grating
10		S.02	http://steel.waste.com/project/S.02	Parrallel Flange Channel
11		S.01	http://steel.waste.com/project/S.01	Universal Beam
12		C.08	http://steel.waste.com/project/C.08	Grating
13		C.01	http://steel.waste.com/project/C.01	Circular Hollow Section
14		C.02	http://steel.waste.com/project/C.02	Circular Hollow Section
15		C.03	http://steel.waste.com/project/C.03	Circular Hollow Section
16		C.04	http://steel.waste.com/project/C.04	Circular Hollow Section
17		C.06	http://steel.waste.com/project/C.06	Parrallel Flange Channel
18		C.05	http://steel.waste.com/project/C.05	Universal Beam
19		M.03	http://steel.waste.com/project/M.03	Grating
20		M.02	http://steel.waste.com/project/M.02	Parrallel Flange Channel
21		M.01	http://steel.waste.com/project/M.01	Universal Beam
22		X.BD	http://steel.waste.com/project/X.BD	
23		X.CD	http://steel.waste.com/project/X.CD	
24		X.MD	http://steel.waste.com/project/X.MD	

Figure A.9: Preprocessed spreadsheet for case study structural data

	E	F	G	H	I	J
1	Parent URI_1	ParentID@RDL	Spec	Cut Length	cutlength_m	Area
2	http://steel.waste.com/rdl/CHS168.3-5.6	CHS168.3-5.6	CHS 168.3x5.6	14 m	14	
3	http://steel.waste.com/rdl/CHS168.3-5.6	CHS168.3-5.6	CHS 168.3x5.6	14 m	14	
4	http://steel.waste.com/rdl/CHS219.1-5.0	CHS219.1-5.0	CHS 219.1x5.0	18 m	18	
5	http://steel.waste.com/rdl/CHS219.1-5.0	CHS219.1-5.0	CHS 219.1x5.0	17 m	17	
6	http://steel.waste.com/rdl/CHS323.9-5.6	CHS323.9-5.6	CHS 323.9x5.6	18 m	18	
7	http://steel.waste.com/rdl/CHS323.9-5.6	CHS323.9-5.6	CHS 323.9x5.6	17 m	17	
8	http://steel.waste.com/rdl/CHS355.69-8.0	CHS355.69-8.0	CHS 355.69x8.0	20 m	20	
9	http://steel.waste.com/rdl/38-5SG	38-5SG	38x5 SG			98 sq m
10	http://steel.waste.com/rdl/PFC260-75-28	PFC260-75-28	PFC 260x75x28	15 m	15	
11	http://steel.waste.com/rdl/UB762-267-197	UB762-267-197	UB 762x267x197	10 m	10	
12	http://steel.waste.com/rdl/38-5SG	38-5SG	38x5 SG			150 sq m
13	http://steel.waste.com/rdl/CHS168.3-5.6	CHS168.3-5.6	CHS 168.3x5.6	14 m	14	
14	http://steel.waste.com/rdl/CHS219.1-5.0	CHS219.1-5.0	CHS 219.1x5.0	18 m	18	
15	http://steel.waste.com/rdl/CHS323.9-5.6	CHS323.9-5.6	CHS 323.9x5.6	18 m	18	
16	http://steel.waste.com/rdl/CHS355.69-8.0	CHS355.69-8.0	CHS 355.69x8.0	20 m	20	
17	http://steel.waste.com/rdl/PFC260-75-28	PFC260-75-28	PFC 260x75x28	14 m	14	
18	http://steel.waste.com/rdl/UB762-267-197	UB762-267-197	UB 762x267x197	15 m	15	
19	http://steel.waste.com/rdl/38-5SG	38-5SG	38x5 SG			70 sq m
20	http://steel.waste.com/rdl/PFC260-75-28	PFC260-75-28	PFC 260x75x28	15 m	15	
21	http://steel.waste.com/rdl/UB762-267-197	UB762-267-197	UB 762x267x197	15 m	15	
22	http://steel.waste.com/rdl/BD	BD	Boat Deck			
23	http://steel.waste.com/rdl/CD	CD	Cellar Deck			
24	http://steel.waste.com/rdl/MD	MD	Main Deck			

Figure A.9: Preprocessed spreadsheet for case study structural data *(Continued)*

Appendix 7: Data Mapping Using dot15926 Editor *(Continued)*

	K	L	M	N	O	P
1	area_sqm	Zone	ParentID@RDL	Parent URI_2	Date	Creator
2		Boat Deck	BD	http://steel.waste.com/rdl/BD	10/5/2017	A.G.
3		Boat Deck	BD	http://steel.waste.com/rdl/BD	10/5/2017	A.G.
4		Boat Deck	BD	http://steel.waste.com/rdl/BD	10/5/2017	A.G.
5		Boat Deck	BD	http://steel.waste.com/rdl/BD	10/5/2017	A.G.
6		Boat Deck	BD	http://steel.waste.com/rdl/BD	10/5/2017	A.G.
7		Boat Deck	BD	http://steel.waste.com/rdl/BD	10/5/2017	A.G.
8		Boat Deck	BD	http://steel.waste.com/rdl/BD	10/5/2017	A.G.
9	98	Boat Deck	BD	http://steel.waste.com/rdl/BD	10/5/2017	A.G.
10		Boat Deck	BD	http://steel.waste.com/rdl/BD	10/5/2017	A.G.
11		Boat Deck	BD	http://steel.waste.com/rdl/BD	10/5/2017	A.G.
12	150	Cellar Deck	CD	http://steel.waste.com/rdl/CD	10/5/2017	A.G.
13		Cellar Deck	CD	http://steel.waste.com/rdl/CD	10/5/2017	A.G.
14		Cellar Deck	CD	http://steel.waste.com/rdl/CD	10/5/2017	A.G.
15		Cellar Deck	CD	http://steel.waste.com/rdl/CD	10/5/2017	A.G.
16		Cellar Deck	CD	http://steel.waste.com/rdl/CD	10/5/2017	A.G.
17		Cellar Deck	CD	http://steel.waste.com/rdl/CD	10/5/2017	A.G.
18		Cellar Deck	CD	http://steel.waste.com/rdl/CD	10/5/2017	A.G.
19	70	Main Deck	MD	http://steel.waste.com/rdl/MD	10/5/2017	A.G.
20		Main Deck	MD	http://steel.waste.com/rdl/MD	10/5/2017	A.G.
21		Main Deck	MD	http://steel.waste.com/rdl/MD	10/5/2017	A.G.
22					10/5/2017	A.G.
23					10/5/2017	A.G.
24					10/5/2017	A.G.

Figure A.9: Preprocessed spreadsheet for case study structural data *(Continued)*

To prepare the point cloud summary for import, the grey coloured columns in Figure A.10 are introduced for preprocessing. Cells A2 and A4 in Figure A.10 contain the namespaces that are used in forming URIs for the data entities. The URIs for platform deck entities are formed by concatenating the namespace in Cell A2 and the unique IDs generated for the them in Column B in Figure A.10. Local IDs in Column E and H of Figure A.10 represent the parent objects (classifiers) of the point cloud entities. The local IDs in Column E are derived using the excel formula:

=CONCAT(LEFT(D2,1),MID(D2,FIND(" ", D2,1)+1,1)) (A7.6)

The local IDs in Column H are all the same because data entities in adjacent Column G are all 3D point clouds. The URIs of the parent objects are formed by concatenating the namespace in Cell A4 (Figure A.10) and the corresponding values of the parent objects in Columns F and I (Figure A.10). Lastly, initial values for platform level and virtual 3D links in Figure A.10 are preserved to import them as annotation properties, and Columns J and K (Figure A.10) contain metadata for the project entities.

Appendix 7: Data Mapping Using dot15926 Editor *(Continued)*

	A	B	C	D	E
1	Namespaces	ID	URI	Platform Level	ParentID@RDL
2	http://steel.waste.com/rdl/	X-BD	http://steel.waste.com/project/X-BD	Boat Deck	BD
3		X-CD	http://steel.waste.com/project/X-CD	Cellar Deck	CD
4	http://steel.waste.com/project/	X-MD	http://steel.waste.com/project/X-MD	Main Deck	MD

	F	G
1	Parent URI_1	Virtual 3D
2	http://steel.waste.com/rdl/BD	C:\Navisworks Data\Delta PP\Revised Truview with Model Superimposed\DATA\BOAT & SUBCELLAR DECK.htm
3	http://steel.waste.com/rdl/CD	C:\Navisworks Data\Delta PP\Revised Truview with Model Superimposed\DATA\CELLAR DECK.htm
4	http://steel.waste.com/rdl/MD	C:\Navisworks Data\Delta PP\Revised Truview with Model Superimposed\DATA\MAIN & UPPER DECK.htm

	H	I	J	K
1	ParentID@RDL	Parent URI_2	Date	Creator
2	3PC	http://steel.waste.com/rdl/3PC	10/5/2017	A.G.
3	3PC	http://steel.waste.com/rdl/3PC	10/5/2017	A.G.
4	3PC	http://steel.waste.com/rdl/3PC	10/5/2017	A.G.

Figure A.10: Preprocessed spreadsheet for point cloud data

Appendix 7: Data Mapping Using dot15926 Editor *(Continued)*

STEP 3: Definition of patterns for data transformation

In defining information patterns, temporal parts, nature or boundary of modelled objects are ignored. This will ensure entity values are stable for lifecycle data storage. To convert the data contents of the spreadsheets to RDF data, information patterns are defined to describe the relations between their entities. Figure A.11 shows the information patterns for the case study data project. All the information patterns have signatures that correspond to the columns of their associated spreadsheets. For the project specific reference data library, a pattern (Figure A.11a) maps the local_ID, specification, part2Type and descrption to the corresponding properties of the object. It also describes a Specialisation entity that captures the relationship between the new entity and its PCA RDL superclass. In addition, the pattern contains two parts for assigning matadata properties (creator and creation date) to the object and entity in the pattern.

To import the structural report spreadsheet data, a pattern (Figure A.11b) maps the local_ID, description, specification and classifiers of the structural object (parentType1) and its location (parentType2) to the appropriate annotation and object properties. Classifiers are used instead of the Part 2 types directly because there may be need to deal with spreasheets with very large data. Using the Part 2 types directly will require each data object to be individually handled. However, using classifiers enable appropriate Part 2 types to be quickly inferred from the project specific RDL.

To map the length and area properties of the structural objects, ISO 15926 template information patterns, 'CutLength' and 'GeometricArea', are adopted. These can be found in the template information pattern source <iip_fullset_20140131_PCA_dm.owl>. Separate parts, M (metre) and M2 (square metre), are created for the templates' units of measurement. In each template, the appropriate unit of measurement and object are matched with the corresponding roles. Also, the length and area values are matched to the corresponding roles. Lastly, separate parts are provided for assigning the metadata properties (creator and creation date) to the object and template entities in the pattern.

Appendix 7: Data Mapping Using dot15926 Editor *(Continued)*

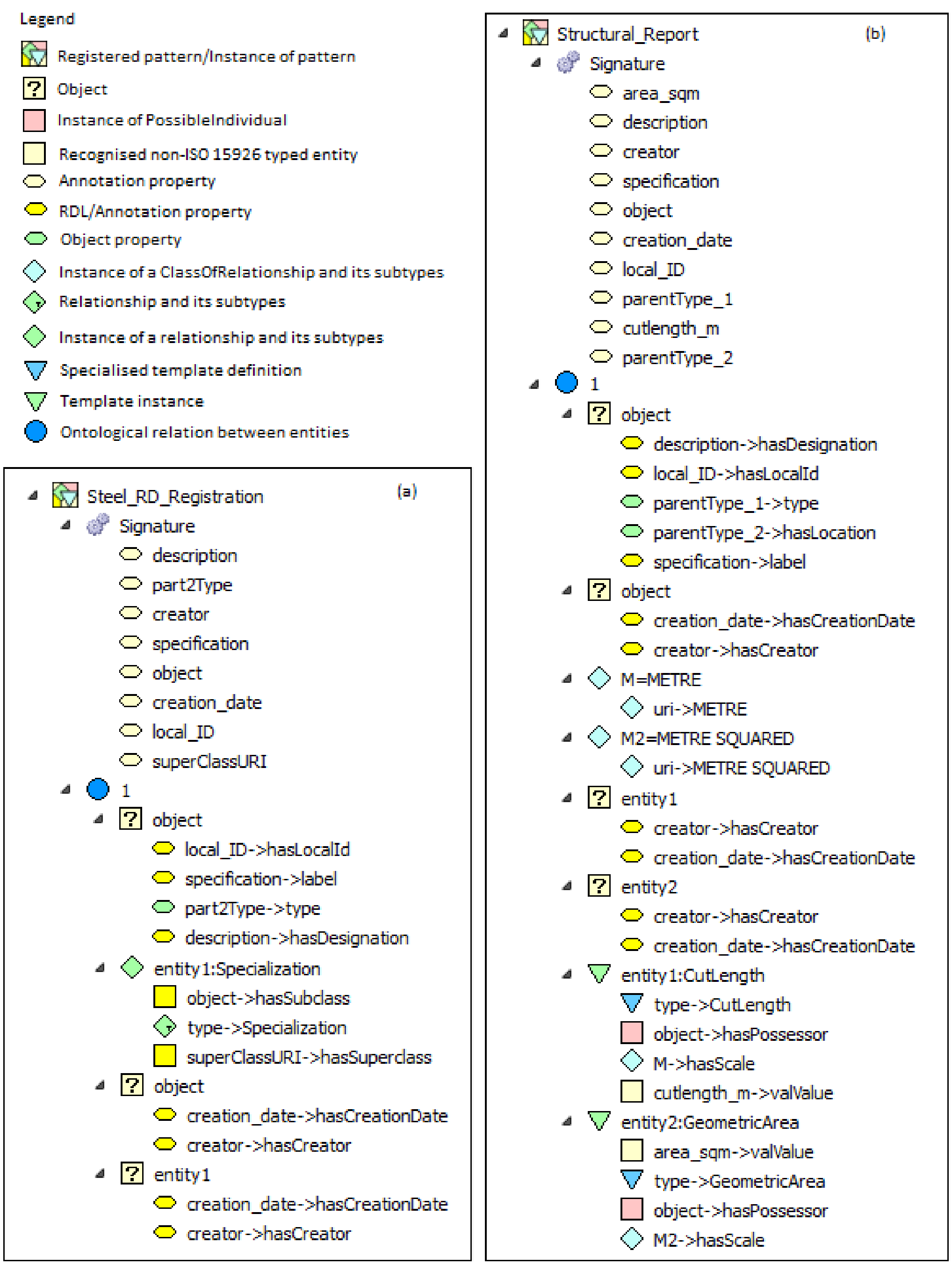


Figure A.11: Case study information patterns

To import data from the point cloud spreadsheet, a pattern (Figure A.11c) maps the local_ID, description, virtual3d link and classifiers of the point cloud object (parentType2) and its location (parentType1) to the appropriate annotation and object properties. Also, a separate part is provided for assigning the metadata properties (creator and creation date) to the object in the pattern.

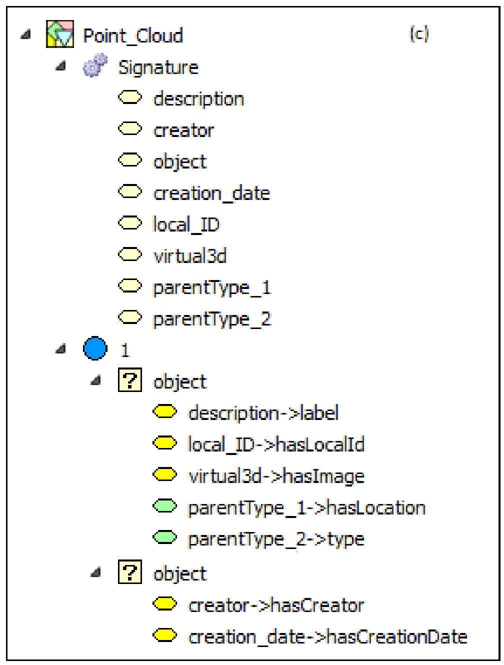
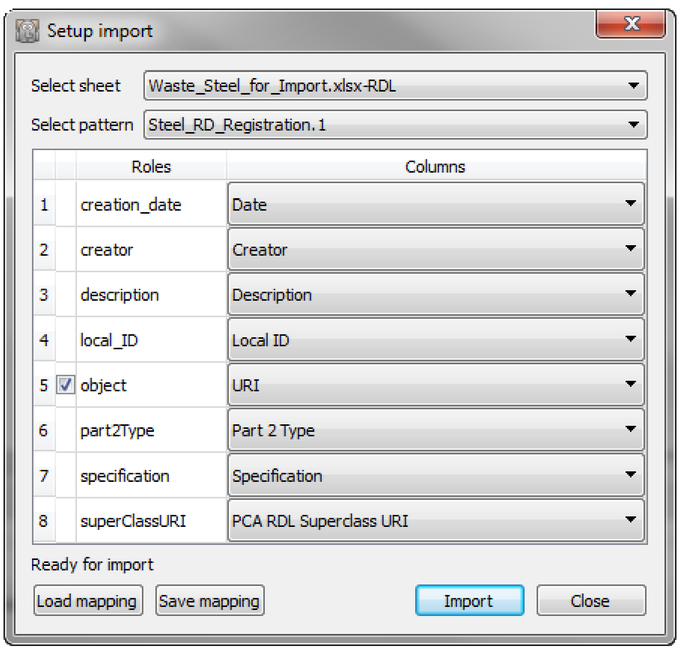


Figure A.11: Case study information patterns *(continued)*

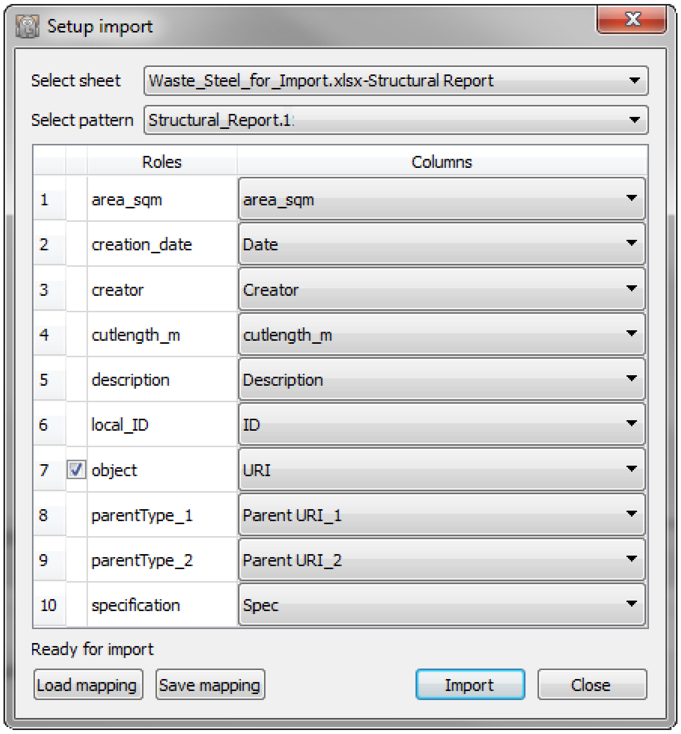
STEP 4: Data transformation to RDF

To import data from the sources, the spreadsheet source should be opened. From the dot15926 project panel, the data file to receive the data (Steel_RDL or Steel_Data) is opened into the data view panel of dot15926 Editor. The visible data file in the data view panel is selected, followed by the 'Build patterns from MS Excel' function from the main menu. A dialog box pops from the latter action. The box allows the spreadsheet columns to be matched to the corresponding entities in the information patterns. Where entities in a role are to be created with URIs recorded in the corresponding spreadsheet column, the associated check boxes are marked. Figure A.12-14 show the corresponding dot15925 Editor mappings for project specific RDL, structural report and 3D point cloud summary respectively. Also shown are the JSON scripts for the mappings. Once the mapping process is completed, the data set generated is saved in dot15926 format. The case study related RDF files (Steel_Data and Steel_RDL) populated in the mapping process are deposited in a triple store for querying.



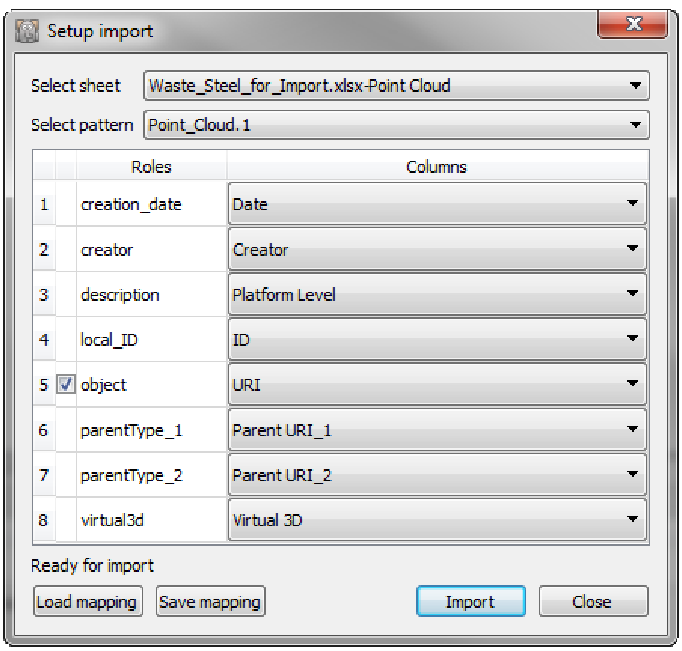

```
{
  "pattern": "RD_Registration",
  "version": 0,
  "option": "1",
  "roles": {
    "description": {
      "name": "Description"
    },
    "part2Type": {
      "name": "Part 2 Type"
    },
    "creator": {
      "name": "Creator"
    },
    "specification": {
      "name": "Specification"
    },
    "object": {
      "name": "URI",
      "build": true
    },
    "creation_date": {
      "name": "Date"
    },
    "local_ID": {
      "name": "Local ID"
    },
    "superClassURI": {
      "name": "PCA RDL Superclass URI"
    }
  }
}
```


Figure A.12: Mapping for project specific reference data




```
{
  "pattern": "Structural_Report",
  "version": 0,
  "option": "2",
  "roles": {
    "area_sqm": {
      "name": "area_sqm"
    },
    "description": {
      "name": "Description "
    },
    "creator": {
      "name": "Creator"
    },
    "specification": {
      "name": "Spec"
    },
    "object": {
      "name": "URI",
      "build": true
    },
    "cutlength_m": {
      "name": "cutlength_m"
    },
    "local_ID": {
      "name": "ID"
    },
    "parentType_1": {
      "name": "Parent URI_1"
    },
    "creation_date": {
      "name": "Date"
    },
    "parentType_2": {
      "name": "Parent URI_2"
    }
  }
}
```


Figure A.13: Mapping for structural report




```
{
  "pattern": "Point_Cloud",
  "version": 0,
  "option": "1",
  "roles": {
    "description": {
      "name": "Platform Level"
    },
    "creator": {
      "name": "Creator"
    },
    "object": {
      "name": "URI",
      "build": true
    },
    "creation_date": {
      "name": "Date"
    },
    "local_ID": {
      "name": "ID"
    },
    "virtual3d": {
      "name": "Virtual 3D"
    },
    "parentType_1": {
      "name": "Parent URI_1"
    },
    "parentType_2": {
      "name": "Parent URI_2"
    }
  }
}
```


Figure A.14: Mapping for 3D scan summary

Appendix 8: Ontology-Based Data Access using Ontop

This section provides information on how to set up virtual RDF repository for querying a relational database source. This is done using Ontop (Calvanese *et al.*, 2017), an open-source Ontology-Based Data Access (OBDA) system. The software supports querying of relational data sources through an ontology used for representing the schema of the data sources, and to which the data in the sources are mapped. Figure A.15 shows the system architecture for OBDA in this research.

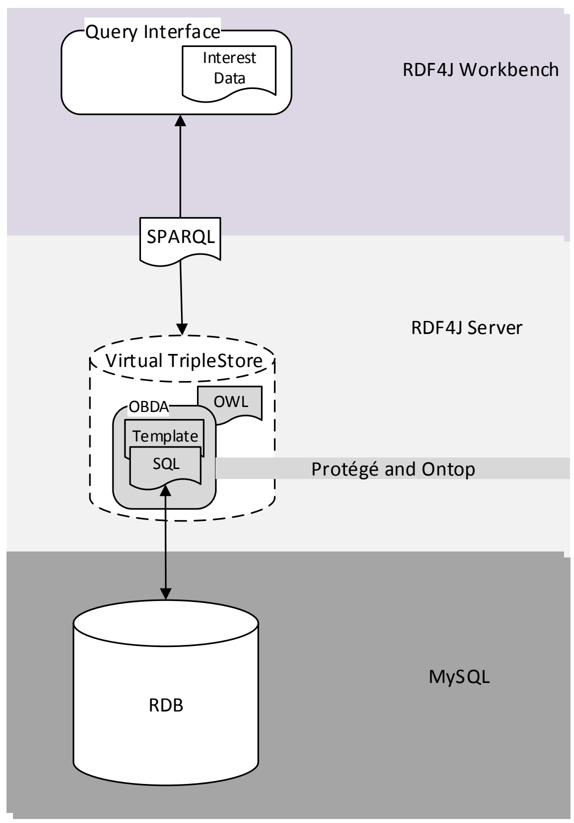


Figure A.15: System architecture for OBDA

The use of Ontop for a virtual repository requires an ontology of the data source in OWL (Motik *et al.*, 2012) and an OBDA model. The OBDA model (file) is composed of

- data source declaration that links a data source to a Uniform Resource Identifier (URI), and to a set of access properties that enable a system to establish connection to the source
- mapping axioms comprising SQL queries over the data source, and RDF triple templates based on the rows returned by the SQL queries.

The mapping axioms of an OBDA model can be generated using R2RML and converted to Ontop's native mapping language using Protégé software (Musen, 2015). Otherwise they can be generated directly using the Ontop plugin in Protégé. Protégé can also be used for generating the ontology of the data source in OWL.

Guidance on software resources for setting up the system architecture in Figure A.15 is as follows:

- [15]*Download and install XAMPP*. XAMMP is a Web Server.
- [16]*Download and Install MySQL database server* – and use XAMPP as the Web Server for running 'phpMyAdmin', the chosen software tool for handling its administration. To activate phpMyAdmin with XAMPP, the start buttons for 'MySQL' and 'Apache' modules in its control panel should be activated. Thereafter, the Admin button for MySQL can be used to connect phpMyAdmin web page - where the database file for this research, ***structural_maintenance.sql*** (Akinyemi, 2018b)*,* can be imported.
- [17]*Download and install Protégé. Also set up the Java Database Connectivity (JDBC) for the relational database to be used (MySQL in this case) and activate the Ontop plugin in the menu.*
- [18]*Guidance on how to generate the ontology of a data source and the corresponding OBDA model is provided in Ontop GitHub Wiki.* The OBDA model and data source ontology (***Maintenance.obda*** and ***Maintenance.owl***) for this research are provided in (Akinyemi, 2018b). Also, the R2RML equivalent of Ontop mapping axiom is provided (***R2RML-mappings***).
- [19]*Download latest version of Ontop pre-packed into RDF4J database server*. From the extracted download folder, copy rdf4j-workbench.war and rdf4j-server.war to XAMPP/webapps folder. Thereafter, use the XAMPP Control Panel to deploy the WAR files. To do this, click the start button for the Tomcat Module in XAMPP

[15] https://www.apachefriends.org/index.html (Accessed 28 September 2018)
[16] https://dev.mysql.com/downloads/mysql/ (Accessed 28 September 2018)
[17] https://github.com/ontop/ontop/wiki/ontopProInstallation (Accessed 28 September 2018)
[18] https://github.com/ontop/ontop/wiki/SimpleHelloWorldTutorial (Accessed 28 September 2018)
[19] https://sourceforge.net/projects/ontop4obda/files/ (Accessed 28 September 2018)

Control Panel. Then click the Admin button to link to Tomcat Web Application Manager. From the latter, RDF4J Workbench webpage can be accessed by using the set login credentials. From this page, a [20]*Virtual RDF Repository can be set up*. The files, ***Maintenance.obda*** and ***Maintenance.owl***, are required for this.

- Lastly, querying can be done via the query interface of RDF4J Workbench. This can be linked to via the Tomcat Web Application Manager.

[20] https://github.com/ontop/ontop/wiki/ObdalibQuestSesameVirtualWB (Accessed 28 September 2018)