University of Bergen

Master Thesis

Examining the development process of a semantic hub in an enterprise environment

Author:
John Fredrik Asphaug

Supervisors:
Andreas Opdahl
Nils Jacob Berland

in the

Semantic and Social Information Systems
Department of Information Science and Media Studies

28th June 2013

University of Bergen

Abstract

Department of Information Science and Media Studies

Examining the development process of a semantic hub in an enterprise environment

by John Fredrik Asphaug

In the resent years there has been developed a couple of semantic hubs in the Norwegian oil and gas industry, and now EPIM is in the process of developing a semantic hub called the LogisticHub. It will be a track and trace system for cargo and containers in the offshore supply chain, utilising Semantic Web technology to share and communicate between the different organisations. The semantic hub as an information system is considered to be a new type of information system, that could benefit from a closer look at its development process. In this thesis I will look at the development process of the LogisticHub, in an attempt to discover the differences in developing a semantic hub compared to the development of other information systems.

Acknowledgements

I would like to thank both my supervisors in this thesis they have both been valuable to get this paper completed, Nils Jacob Berland for providing me with connections and information from within the Norwegian oil and gas industry, and Andreas Opdahl with guidance on taking difficult choices surrounding the research and its findings. I would like to thank Statoil for inviting me to follow them in this project, and for taking the time to have several meetings with me, discussing the project providing me with valuable information. I would also like to thank ConocoPhilips and EPIM for letting me interview one of their representatives. Additionally I would also like to Thomas Olsen for proofreading my thesis paper, removing some of the embarrassing typos. Last, but not least I would like to thank my beloved Silje Bjønnum that had to suffer with me through this thesis and gave me lot of support

Contents

List of Figures

List of Abbreviations

BPMN Business Process Management Notation

CCU Cargo Carrying Unit

CRUD CREATE READ UPDATE DELETE

EPIM E&P Information Management Association

OIM Offshore Installation Manager

RFID Radio Frequency Identification

URI Uniform Resource Identifiers

URL Uniform Resource Locator

Chapter 1

Introduction

1.1 Norwegian Oil and Gas

In the last 40 years the oil and gas development on the Norwegian continental shelf has become the base of the Norwegian economy. Today there are more than 500 active production licenses with high exploration activity and more than 80 fields in production. 40 of these operators have an annual budget close to 200 billion Norwegian Krone. 80% of this budget is used on products and services delivered by suppliers. For a long time there has been little consideration to the costs in the Norwegian oil and gas industry, but in recent years this has become more of a focus, creating demands for lowering costs, and increasing efficiency and productivity. One way to do this, has been to look at ways of improving the offshore supply chain. As the result a track and tracing solution has been proposed as a step to reduce costs in the offshore supply chain, it has been estimated to save hundreds of millions of Norwegian Krone(The Norwegian Oil and Gas Association, 2012, (PD3)).

1.2 E&P Information management Association (EPIM)

EPIM is a joint venture owned and managed by the oil and gas operators on the Norwegian Continental Shelf. EPIM works with the objective to utilize and make available information technology solutions that facilitate the best possible flow of information between its users. (E&P Information Management Association, 2012). They aim to create a large knowledge base by creating several semantic hubs one at the time to fill different areas of the oil and gas industry. The latest project is the LogisticHub(IN3).

1.3 Semantic Hub

A semantic hub is an information system that aims to become a network of information. Semantic information can be seen as explicit knowledge where all meanings are expressed clearly and with nothing left to be implied. Knowledge can also be seen as accumulated information, and thus knowledge contains the ability to use information effectively in a specific context. In this case the semantic hub is a knowledge base that has an information model at its core. The information model provides the explicit meaning, to knowledge in the semantic hub. The information model in a semantic hub is more commonly referred to as a vocabulary or ontology. The ontology is a set of concepts and terms that give meaning, to the data stored in the semantic hub. By adding meaning to the data, computer agents can access the data automatically, and know not only what the data is, but also what it represents (Avison & Fitzgerald, 2006; Berners-Lee, Hendler & Lassila, 2001).

1.4 LogisticHub

In 2010 a guideline was published(Norwegian Oil and Gas Association, 2010) for using Radio Frequency Identification(RFID) technology in the Norwegian oil and gas industry. These guidelines highlight everything from the deployment of RFID technology, technical specifications, to general principles, architecture and integration. LogisticHub is the implementation of the RFID tracking system. LogisticHub's main aim is to coordinate and collect tracking information of cargo carrying units CCU. LogisticHub communicates data between users in events. Events are triggered, and information surrounding the events are collected and stored in the semantic hub. An event contains five main attributes where, when, what, who and why events triggered.

- Who - triggered the event. This will be the organisation that created the event.

- What - type of CCU does this event concern. In phase two of LogisticHub it is envisioned that this attribute will also list what items are stored within the container.
- Where - was the event triggered. This will be a geographical location. Most likely a GLN location number with an longitude latitude coordinate.
- When - was the event triggered. A timestamp of when the event was triggered in an office time format yyyy.mm.dd.hh.mm.ss.
- Why - is the event triggered is suppose to link the event to business logic. Giving a business explanation to why this event was triggered.

1.5 Goal and Research Question

This thesis will try to examine the process of developing a semantic hub in a enterprise environment. By following the development process of the semantic hub LogisticHub, I want to examine `how the process of developing a semantic hub compares to the process of developing other information systems?`

To get a deeper understanding of the development process of a semantic hub, one must take into consideration multiple variables. I have decided to explore the process by looking at several different sub-goals as a way of reaching the main goal.

One of the earlies stages of development is to identify possible benefits from the new system, and its desired outcome. The desired benefits and desired outcome are often used during the development process as the long term goals for the new system, and might be the factors that decide what type of system to develop. In order to better understand the development process, and what triggered the development of the semantic hub. I will use the case of the LogisticHub to examine, `which benefits are seen by the organisation as desired outcomes for the semantic hub?`

The nature of a hub is to be a centre of activity. The semantic hub will be the centre of the activities in the business processes surrounding the offshore supply chain. The different user organisations will need to integrate the semantic hub with their existing systems to fully utilise it. With the large number of organisation that will be using the semantic hub, I want to examine `how do the organisations intend to use the semantic hub?` This as a step in examining if the uses intended for the semantic hub effects how the development process of the semantic hub will be conducted. Through the development of a system, the developers use system development methods to guide them in their choices of development techniques that are appropriate at the different stages in the development process. The development techniques are used to ensure that each development stage is done thoroughly, and thus the whole development process. The development techniques are considered to play an influential role in how the information system gets developed. I want to examine how some of the more commonly used development techniques are used in the development of a semantic hub to examine `which development techniques could fit with the development process of a semantic hub?`(Avison & Fitzgerald, 2006)

All types of information systems are constructed on the basis of different technologies. Technologies provides technical solutions for the the information systems to reach the desired long term goals. When building a semantic system like a semantic hub you are dependent on Semantic Web technologies. Even though the Semantic Web technologies are not considered as a new technology in the academic world, they are still considered to be a new technology in a enterprise environment. I want to test some of the different Semantic Web technologies used in a semantic hub, as a way of examining `how the technologies used in a semantic hub have impact on the development process?`

I ended up with this main research question:

`How is developing a semantic hub different from developing other generic information systems?`

An these sub question that I will attempt to answer in this thesis:

RQ1 `Which benefits are seen by the organisation as desired outcomes from the semantic hub?`

RQ2 `How do the organisations are intending to use the semantic hub?`

RQ3 `Which development techniques could fit with the development process of a semantic hub?`

RQ4 `How do the technologies used in a semantic hub impact on the development process?`

Chapter 2

Theory

2.1 Interoperability with Semantic Web

It all started with the internet and its creator Tim Berners-Lee. He envisioned and created a web of documents we now call the internet. The internet, or World Wide Web is designed for human consumptions of information. Even if all the information on the web is machine readable, does not make the information machine understandable (Lassila, Swick, Wide & Consortium, 1999). In 1997, Tim Berners-Lee had an extended vision for the world wide web. He wanted to turn the web into a web of meaning, to set expressive meaning to all the data and information flowing around on the great World Wide Web (Veltman, 2001; Berners-Lee et al., 2001). He called this vision the Semantic Web. The basic goal of the semantic web is to make the web machine understandable.

That is, to take the enormous amounts of information located on millions of different web pages, and make these processable by machines. This so that the machines can a better aid us in getting the information we need. Berners-Lee et al. (2001) builds the semantic web on already existing concepts, and technology that is already in use and has proven its value. With the use of hyperlinks, also known as Uniform Resource Identifier (URI), data is tagged with its meaning.

These representations of metadata are constructed as statements, and each statement consists of a triple. The triple is modelled after the natural language and contains a subject, predicate and object. The most common representation of metadata in the semantic web is through the use of the Semantic Web technology called the Resource Description Framework.

2.2 Resource Description Framework (RDF)

Resource Description Framework is the basic building block for processing of metadata. RDF provide interoperability between applications that exchange machine understandable information on the web. RDF is supposed to define a domain neutral mechanism for describing resources without making assumptions about any particular domains. RDF is structured in classes called schemas. These are structured in much the same way as other object-oriented programming and modelling systems. These schemas have an hierarchy structure that offers expandability through subclass improvement. This enables us to reuse schemas, making it easier since we only need to slightly alter or refine the old schema to fit our domain (Berners-Lee et al., 2001).

RDF is represented in a model by naming properties and property values. RDF properties follow the tradition of attribute-value pairs; it can also represent relationships between the resource and the RDF model. Two RDF models are seen as equal if and only if their data model representations are the one and the same (Lassila et al., 1999).

A RDF model is composed of three types of objects: resources, properties and statements. Resources are all the things described with RDF. The properties are used to describe a specific aspect, characteristic, attribute or relation of the resources. A statement is also better known as a triple. These triples are structured similar to the natural language with a subject and object and a predicate to link the relation between them. The subject and predicate are always represented as a resource by an Universal Resource Identifier. This si more commonly known as a hyperlink or URL. URL is an subset of URI and is the most commonly used URI. The object of the statement is either a resource or a literal. A literal is a primitive data type like a string (Berners-Lee et al., 2001; Thi, Thuy, Lee, Lee & Jeong, 2007; Dave Beckett & MacBride, 2004).

In RDF it is allowed for objects and values to be interchanged. Making it possible for the object to be the value and thus giving the option to connect to edges in a RDF graph. RDF also supports a form of reification, which any RDF were any statement can be object or value of a triple, making it possible to have the graphs nested as well (Decker et al., 2000)

RDF is built on a open world assumption which makes it possible for anybody to make a statement about resources. On can therefore never assume in general that there is complete information about any resource available.

2.3 Named graph theory

One view of the semantic web is that its a collection of linked RDF documents, each containing RDF graphs. The RDF specifications (Dave Beckett & MacBride, 2004) have descriptions and explanations for the meaning of one graph and how one may merge a set of graphs into one, but it does not provide a mechanism for talking about the graphs themselves, or the relationship between the graphs. Carroll, Bizer, Hayes and Stickler (2005a) propose a new simple variation of RDF called `named graphs`. A named graph simply extends a RDF graph by giving it a unique name in the form of a URI reference. A named graph can either occur in the graph itself, in other graphs. Different RDF graphs may share URI references, but not blank nodes. Another way to view named graphs are as a reformulation of quads statements. Where the fourth elements semantics and syntactic properties are more clearly distinguished. In addition to the relationship to the RDF triples abstract syntax and semantics become clearer. Named graphs are intentionally built as a small step on top of RDF and OWL recommendations. Thus allowing the use of named graphs with tools built as implementing those recommendations, in a backward compatible way (Carroll et al., 2005a).

Since RDF and OWL function under the open world assumption, it is clear that the description of the resources are considers to be open ended. Tangible information as web documents, either web pages or RDF graphs can thus be identified and rigidly named, making this name uniquely identify the resource. This is utilized by making the name attach to the graph rigid. Thus mapping between named and graphs fixes the graphs corresponding to a named in a rigid, non extensible way (Carroll et al., 2005a).

Carroll et al. (2005a) claims that by following definition of a reified statement as a single RDF statement described and identified by a URI reference(Hayes, 2004), it becomes natural to think of named graphs as

W3C has adopted a slightly modified version of named graphs. The main difference from the work of Carroll et al. (2005a) is by adding one unnamed default graph. This provides the backward compatibility with RDF without named graphs that allows the named graph functionality of SPARQL to be optional. Carroll et al. (2005a) argues this that this slight modification will reintroduce some of the difficulties that named graphs address. As an example: when merging both default graphs and named graphs from different repositories, while trying to maintain provenance information. This is specially a problem when trying to serialise documents of RDF datasets using syntaxes like TriX or TriG which serialise multiple graphs into a single document.

2.3.1 TriX & TriG

To serialise named graphs, Carroll et al. (2005a) presents two different syntaxes, TriX and TriG. TriX is based on XML and tries to have a basic syntax corresponding closely to the RDF abstract syntax, using different XML elements to define the graphs and triples.

TriG serves as a compact and easily readable alternative to TriX. TriG is based on N3, built as an extension Turtle (David Beckett & Berners-Lee, 2008) using { and } to group triples into multiple graphs, and to precede each by name of that graph. In Figure 2.1 the first named graph called `:G1` states information about itself and the second graph `:G2` states information about the first graph.

```
:G1 {  _:Monica ex:name "Monica Murphy" .
_:Monica ex:email <mailto:monica@murphy.org> .
:G1 pr:dialloedUsage pr:Marketing }

:G2 { :G1 ex:author :Chris .
:G1 ex:date "2003-09-03"^^xsd:data }
```

FIGURE 2.1: Example of named graph written in TriG(Carroll et al., 2005a)

TriG has syntactical compatibilities with N3, but some of the N3 attributes are not allowed.

- Blank nodes can not be shared between graphs
- Each graph should have a named URI reference

- A formula for a graph cannot be embedded within another graph as a node

These differences entail that basic syntactical operations like comparing two N3 documents for differences in their abstract syntax, or signing the abstract form of an N3 document, differ significantly, and are substantially more difficult than the corresponding operation on a simple RDF graph.

2.4 SPARQL query language

RDF represents data on the web as a labelled graph. It is used very often to represent a very wide range of information and enabling the information to be integrated over the disparate information sources (Garlik, Seaborne & Prud'hommeaux, 2012). SPARQL was introduced in 2004 and is now the recommended query language by W3C. SPARQL is basically a graph matching query language. Where you have a specific data source, and a query consisting of a pattern, that is matched against patterns in the data source. The values matched from the query pattern are returned as the query result (Pérez, Arenas & Gutierrez, 2006). Triple patterns in SPARQL are like RDF triples except each of the subjects, predicates and objects can be represented as a variable. A simple SPARQL query consist of two parts: a `SELECT` clause an a `WHERE` clause. The `SELECT` clause identifies variables to appear in the result and the `WHERE` clause provides the graph pattern to be matched against a data graph. The triple pattern may consist of a single triple or more complex patterns included that might include extended features like `FILTER`, `UNION` etc. The example in Figure 2.2 contains one triple with `?name` as variable in the object position.

```
SELECT ?name
WHERE
{
<http://example.org/john> <http://xmlns.com/foaf/0.1/name> ?name .
}
```

FIGURE 2.2: Simple SPARQL query

2.4.1 SPARQL 1.1 UPDATE

SPARQL 1.1 UPDATE was introduced making it possible to update a triplestore through a SPARQL endpoint. Similar to other query languages it now has the option to either insert or delete data in the triplestore through the SPARQL endpoint. INSERT DATA adds triples into a graph store and DELETE DATA removes triples from a graph store. SPARQL 1.1 UPDATE has support for deleting and inserting triples from a graph store based on bindings connected to a WHERE clause. By first using DELETE and then INSERT SPARQL provides an option for UPDATE of triples in a graph store. (Garlik et al., 2012).

2.5 Ontology

A major purpose of the semantic web is to link data together. To achieve this there needs to be a common understanding of the meaning of things. For the data to be linked together, and for the machines to understand that people are talking about the same things, people need to agree on a common vocabulary for describing the meaning, which they then can tag in their documents on the web with this meaning. Hebeler, Fisher, Blace and Perez-Lopez (2011) defines an ontology as a predefined, reserved vocabulary of terms to define concepts and the relationship between them for a specific area or domain. The term ontology is borrowed form philosophy and refers to the science of describing the types of entities in the world and how they are related. Search engines use ontologies to find pages that have words that are syntactically different, but semantically similar (Decker et al., 2000). Ontologies provides a shared and common understanding of a domain that can be communicated between people and across application systems (Jung, Park, Cha & Jo, 2008). This strengthen the information that already exists, by connecting it to a vocabulary (Kinsella, Bojars, Harth, Breslin & Decker, 2008). Hendler (2001) defines ontology *"as a set of knowledge terms, including the vocabulary, the semantic interconnection, and some simple rules of inference and logic about a particular topic"*. Ontologies combined with logic and rules of inference make it possible for a machine to reason on the information in the model and thus increasing the knowledge representation. And with the support, for chaining and nesting RDF graphs one can create enormously large data sets that machines can navigate and draw information from by reasoning. That gives the user better and more accurate information according to the query they might have had (Hendler, 2001; Decker et al., 2000). This enables reuse of domain knowledge and has been one of the major driving forces for developing ontologies and using the semantic web technologies (Noy & Mcguinness, 2000).

Classes are the focus of ontologies. They describe and represent the domain of the ontology. An example might be a class of cars. Specific types of car are instances of the car class, but you also might have subclasses. Subclasses of cars might be SUV or covertible. These subclasses specifiy more specific concepts in the ontology. In this case a more specific type of car.

2.5.1 Building an ontology

When building a ontology one need to understand for what purpose it is inteded to be used. Creating the need to identify what the ontology has as an intended goal (Uschold, 1996). Gruninger and Fox (1995) defines the goal for any given ontology to be an agreed upon shared understanding terminology for the objects in the ontology. Ergo the users of an ontology must agree on the purpose and the use. Uschold (1996) argues that the type of ontology you are building will shape how you build it. He therefore recommends that one starts ontology building by defining three key dimensions:

Formality: Describing the degree of formality in which the ontology is created and how the meaning is specified. Should the ontology be specified loosely and based only on natural language, or more rigorously by defining terms in formal semantics, theorems and proofs of soundness and completeness.

Purpose: Simply states what the intended use of the ontology is. Is it suppose to be used as communication between people, interoperability among systems, or other benefits related to system engineering.

Subject matter: Comes back to what domain should the ontology describe. What is it the ontology will describe in terms and concepts.

Exchanging data across systems, and at the same time providing readable data that people with less system technical background can undetstand. Having now defined both formality and purpose of the ontology one needs to limit the domain in which the ontology will cover.

Noy and Mcguinness (2000) presents a very good starting point in developing your ontology, by suggesting a couple of questions that will aid researchers in limiting the scope and domain.

- What is the domain the ontology is going to cover?
- What will the ontology be used for?
- What types of questions in the ontology should provide answers?
- Who will be using and maintaining the ontology?

They further state that ontology building is an iterative process and therefore the answers to these questions will change over time. These questions overlap with the dimension given by Uschold (1996), but together give a more wholesome picture of how one should build the ontology.

To further determine the scope of the ontology, Gruninger and Fox (1995) introduce competency questions. These are questions that are written down as a list of questions that the knowledge base of the ontology should be able to answer. Uschold and King (1995) support this recommendation and further claims that the competency questions play a role of reference document in a comprehensive methodology. For the researcher to be able to construct these competency questions they will need to familiarise themselves with the domain that the ontology aims to conceptualise.

Even thought these question will be used later to test the ontology, they are only a sketch and do not have to be very extensive (Noy & Mcguinness, 2000). They are also iterative in the way that they constantly change as the ontology takes shape and as scope and domain changes become clear. Noy and Mcguinness (2000) recommends reviewing other ontologies to see if there exist one in your domain that might be possible to refine and extend to your needs.

The next step is to define classes and hierarchy of the ontology followed by the relations between them (Noy & Mcguinness, 2000; Kapoor & Sharma, 2010). When starting to write up the class hierarchy of the ontology three ways are proposed. When wrting using the `Top down` method, one start by first creating the most general of concepts in the domain and then go down to creating more specialised concepts. As example a class *Drink* could be a more general concept of *Coffee*. Then *Coffee* can have subclasses as *Espresso* and *Macchiato*. Another way to start is `bottom down` where the development process starts at defining the most specific concepts first and then work your way up to the more general concepts. Where as example *Pink lady* might be the lowest form of concept registered in a fruit ontology. Its natural superclass might be *Apple*. The third method is called `combination` by Noy and Mcguinness (2000) which have a close resemblance to a `middle-out` method purposed by Uschold (1996). In this method you pick your concepts and start defining subclasses or superclasses for that concepts. Even though none of these approaches are better than the other, the combination method would be best suited for most ontology builders, since the middle concepts are most often the descriptive concepts of the domain (Noy & Mcguinness, 2000).

2.6 Access Control in the Semantic Web

An important part of interoperability and sharing data is to only share that which is intended to share. Access control mechanism usually is composed of two parts authentication, identifying the user and authorisation giving the user access to a system object like a web service (Belchior, Schwabe & Parreiras, 2012). This is not enough to handle access control on the Semantic Web. The semantic hub may store data that is owned by many different organisations and not all this data should be accessible by all. Therefore it is very important that access to the data stored in the semantic hub is controlled down to a fine grained level of the statements stored in it. There is little support in today's existing triplestores for access control. The most common way to handle access control is to limit access to the triplestore. When the access is needed to the data you get access to all data stored in the triplestore. To be able to limit access to specific data the access control needs to support restrictions down to a to triple level and not only at repository level (Flouris, Fundulaki, Michou & Antoniou, 2010).

This is important since the triples statements themselves are not annotated with accessibility information, but an enforcement mechanism is implemented through the injection of the permissions in the query, making sure that only accessible triples will be obtained. Many access control frameworks use access control permissions and access control policies. The access control permission are used to explicitly set certain triples in an RDF to be accessible or inaccessible. An access control policy includes a set of access control permissions and information to determine whether a triple is accessible. When a triple statement is not associated with any permission, or when conflicts arise in permissions of a statement, where one gives access and one restricts, permissions is denied. It is most common to set the initiated state of all statements as not accessible. To not lose any quality of the data the access control need to be on the triple lowest form of data to protect. Giving protection at a pattern matching level to support inference and also support explicit propagation of RDF authorisations, so that when an authorisation is specified for an upper concept, the same access authorisation is also applied to all lower concepts over the ontology hierarchy by inheritance (Gabillon & Letouzey, 2010; Park, 2008; Kim, Jung & Park, 2008; Costabello, Villata, Delaforge & Gandon, 2012).

2.6.1 Access control broker

Costabello, Villata et al. (2012) propose an access control framework to handle access rights to RDF data. They claim the main advantage for this model is that it can easily be implemented into any SPARQL endpoint and thus be used on any triplestore. It serves as an extra layer on top of the triplestore as an access broker handling all traffic to and from the triplestore.

The access control framework uses an access broker that takes the users query and rewrites the query to match the users access rights. This access control model adopts the granularity of named graphs, and thus supporting fine grained access control policies down to the triple level (Carroll et al., 2005a). The choice to rely on named graphs instead of documents is that one document can serialise several named graphs, one named graph can extend over several documents, and not all graphs come from documents. The access rights are grounded on the S4AC ontology (Villata, Delaforge & Gandon, 2011). Access control is handled by access policies, an access policy is shown in Figure 2.3. The access policies contains a set of named graphs that the users get access to if he fulfils the demands from the access control condition. Each access policy contains an access condition set which contains a set of access conditions that need to be fulfilled in order for the access to be satisfied. These access conditions are written in form on ASK queries(Figure 2.4). If an access condition returns true access is given, if false access is rejected. An access condition can be set to either be disjunctive or conjunctive. If the set is conjunctive all access conditions must be true in order to give the user access. If the set of conditions are disjunctive, only one of the access conditions must be true. The access policy also needs to be given an access privilege. Costabello, Villata et al. (2012) divide access privilege into Create, Read, Update and Delete (CRUD) making it a more fine grained access control beyond read/write privileges. To show which graphs are protected by the access policy, named graphs are mapped with `s4ac:appliesTo` either by writing the URI of one or more named graphs or by giving query that returns a set of named graphs. The named graphs get added to the SPARQL query using FROM NAMED (Figure 2.5). The query gets executed it only returns results from the query.

```
:ap1 a s4ac:AccessPolicy ;
     dc:title "Access to inspection for Statoil " ;
     s4ac:appliesTo :G2 ;
     s4ac:hasAccessPrivilege [ a s4ac:Read ] ;
     s4ac:hasAccessConditionSet _:acs2 .
```

FIGURE 2.3: Example of an access policy

To evaluate these access policies, a context graph is created from information about the user (Figure 2.7). This graph is created both from user information stored in the triplestore and context information gathered from the users device. An example of this might be user `John` who wants to query his own name to find his URI (Figure 2.6). The

```
:acs1 a s4ac:AccessConditionSet ;
       a s4ac:ConjunctiveAccessConditionSet ;
       s4ac:hasAccessCondition [
           a s4ac:AccessCondition ;
           s4ac:hasQueryAsk """
              PREFIX lho: <http://logistichub.org/>
               PREFIX foaf: <http://xmlns.com/foaf/0.1/>
               PREFIX prissma: <http://ns.inria.fr/prissma/v2#>
               ASK {
                    ?context a prissma:Context ;.
                    lho:employedAt lho:Statoil.
               }
           """
       ] .
```

FIGURE 2.4: Example of an access condition set with one access condition

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX lho:  <http://logistichub.org/>

FROM NAMED lho:G2

SELECT *
WHERE {?uri foaf:name "John"}
```

FIGURE 2.5: Example of SELECT query with FROM NAMED

```
PREFIX foaf:<http:xmlns.com/foaf/0.1/>
SELECT ?uri
WHERE {?uri foaf:name "John"}
```

FIGURE 2.6: Example query for URI by name

```
  ag: <http://accessgraph.org/>
 ag:john {
   lho:john  a  lho:Person , prissma:Context , rdfs:Resource ;
       lho:employedAt lho:Statoil ;
       lho:typeOfAction lho:typeOfAction ;
       prissma:user lho:base ;
       foaf:mail "john@statoil.com" ;
       foaf:name "John" ;
       foaf:phone "55112233" .
       }
```

FIGURE 2.7: Example of a context access graph

query will be sent to the access broker. The access broker will receive both the query and `John's` context access graph(Figure 2.6 and 2.7.

The access broker will query the users context graph by binding it to the access conditions ASK queries. These queries will return either true or false to the access policy. The access policy if true will then return all named graphs `John` has access to, in this case graph lho:G2 (Figure 6.1). Then add those graphs to the query in form of a FROM NAMED lho:G2 (Figure 2.5) and execute returning only results in which John has access in this example `lho:john foaf:name "John"`.

2.7 Information System

Avison and Fitzgerald (2006) presents a definition of an information system as: *A system which assembles stores, processes and delivers information relevant to an organisation*

```
PREFIX lho: <http://logistichub.org/>
              PREFIX foaf: <http://xmlns.com/foaf/0.1/>
              PREFIX prissma: <http://ns.inria.fr/prissma/v2#>
              ASK {
                   ?context a prissma:Context;.
                   lho:employedAt lho:Statoil.
              }
              BINDINGS ?context {ag:john}
```

FIGURE 2.8: Example of access condition ASK query with BINDINGS

(or to society), in such a way that the information is accessible and useful to those who wish to use it, including managers, staff, clients and citizens. An information system is a human activity (social) system which may or may not involve the use of computer systems. I an organisational environment these information systems are most likely computer-based information systems. Computers can process data at high speeds with very good accuracy, and provide this as information to some useful purpose. The difference between data and information is that data is not interpreted, whereas information has meaning and use to a particular recipient in a particular context, and can often be used for decisions making. Information on the other hand, comes from selecting data and presenting in a way that makes it useful to the receiver (Avison & Fitzgerald, 2006).

The information systems are related to each other. Information systems often have sub systems within it. Sometimes systems can be seen as the collection of sub systems that together become the super information system. The system part of an information system is often seen as a set of interacting components, consiting of: people that use the system, objects computer hardware devices, user interfaces, and procedures like business processes, business rules.

Systems are also often designed to provide information to the users to aid with decision making. Information needs to be presented at the right time, with the appropriate level of detail and of sufficient accuracy to be used by the recipient. There are many different ways of developing information systems.

2.8 Information System Development

Avison and Fitzgerald (2006) define information system development as: *A collection of procedures, techniques, tools and documentation aids which will help the systems developer in their efforts to implement a new information system. A methodology will consist of phases, themselves consists of sub phases, which will guide the systems developer in their choices of the techniques that might be appropriate at each stage in the project and also help them plan, manage, control and evaluate the information systems project.*

The definition of information system development claims that the methodology consists of several phases. Although there are many varieties of system development methodologies they do have a couple of basic phases (Avison & Fitzgerald, 2006):

Feasibility Study looks at the present system trying to find requirements for the new system. The study may consist of reviewing the present system, the requirements that this system was intended to meet, problems it had in meeting these requirements, new requirements that might have revealed themselves since it was first implemented, and a brief investigation into alternative solutions. From these, a recommended solution suggesting, pointing to what the organisations want with the new system. The feasibility study provides a basis to understand why the organisation want the new system.

System investigation is a thorough investigation of the system's application area, going deeper than the feasibility study. It attempts to determine how the organisation intends to use the new system. This is done by observing the use of existing systems, interviewing users to provide their perceptions, questionnaires to collect information for a large group of users, and old records and documentation to highlight the problem. This stage requires a great deal of skill, and by using several of the methods, the results can be cross checked and verified with each other.

System analysis uses the facts from the investigation to analyse the system, in an attempt to understand why the system is developed, and indicate how things might be improved with the new system. The analysis particularly emphasize the the need to establish the requirements for the new system.

System design is taking the new facts revealed in the previous phases and using these to either adopt or change the initial design. Both the computer and manual parts of the system are designed in this phase. The design documentation will contain a set of details about input data and how the data is captured, outputs from the system, processes that need handling, security and backup, and systems testing and implementation plans.

Implementation is the part where the actual information system gets coded and fitted to the organisation. All aspects of the system needs to be proven before it is cutover, else failure will cause lack of confidence in the new system and future computer applications. Quality control is an important part of this phase. The users, as well as the analysts, need to find the system satisfactory. Giving staff good education and training in the new system will make the transition from the old to the new system much easier. The education and training of staff will also more likely give the users a way to better cope with a new systems approach making it easier to become comfortable with the new methods in it. The security procedures for the new system needs to be tested to ensure that there are no unauthorised access and recoveries possible. Once all these things are in place, the new system can either be cutover overnight or parts of the new system can be implemented in turns forming a phased implementation.

Review and maintenance is the final phase in the development and occurs when the new system is operational. There will always be things that need changing, so staff will need to perform maintenance on the system to ensure that it runs continuously and efficiently. Changes will happen for different reasons, some due to organisational or environmental changes, some because of technological advances, and some due to extra needs added to the system. There will also at some point be done a review of the new system, with the aim to confirm the requirements set at the feasibility study phase, and that the cost have not exceeded that which was predicted. There are often conflicts between the operational system and the requirements laid out in the feasibility study, this sometimes leads to a new look at the system to either enhance it or develop another new one, thus starting the whole cycle again.

2.9 System Development Techniques

System development techniques are seen as the way to do a particular activity in the information system development process. The different techniques uses one or more tools to perform the development. The techniques are seen as aids for carrying out tasks and help in capturing the problems from several angles. As an example these development techniques might help to evaluate the costs and benefits of different solutions and methods, to provide detailed designs needed to develop a computer application. By using a mix of different techniques, developers can be provided with a wider understanding of the problem to solve. It is generally seen as beneficial to use system development techniques, but some argue that by using these techniques, one might restrict the understanding by framing the way people think about a project. This will not give them the full understanding, and create a hindrance for the best possible analysis, design and implementation. With effective use of information system development techniques, the information system development attempts to provide effective use of information system technologies. Another part of information system development methodologies is to balance the technical aspects with the behavioural aspects(Avison & Fitzgerald, 2006).

2.9.1 Unified Modelling Language (UML)

UML is an openly developed system development technique to aid and model information system development. It is used to specify and visualise, construct, modify and document an artefact that is under development. This is to help streamline the development of software systems, by giving developers the possibility to visualise and document these models, including their structure and design. (Object Management Group, 2013)

2.9.1.1 UML class diagrams

The UML class diagram is one of the central diagrams in the UML notations. This diagram display the classes and the relations between them in your application. You create the other diagrams primarily to refine your understanding of what is going on in the class diagram. The UML class diagram is seen as a structural diagram that shows the static structure of the information system. They are often used to validate the system between different teams, like business analysts and system architects. The main purpose for the diagram is to model the different types within the system (Fowler, 2004).

2.9.2 Business Process Management Notation (BPMN)

Business Process Management Notation (BPMN) is a development technique used to model the flow in a system. It was to created as a common notation to coordinate the sequences of processes and messages that flow between different processes and in a related set of activities. It aims to make the business processes understood by both the business users, who creates initial draft of the business processes and by the technician that will implement the system. A common objective with BPMN is to model collaborative business processes. The model depicts the interaction between two or more business entities. Where the interactions are shown as sequences of activities, and the message exchange patterns between the participants(Object Management Group, 2011).

Chapter 3

Methodology

This chapter will describe the research method that has been deployed in this thesis. First decribing the research method chosen for this thesis paper.

3.1 Research Method

Going deep into the context is a characteristic of a qualitative research method and the interpretive perspective on the conduct of research. Qualitative research strategies emphasize an interpretive approach that uses data to both pose and resolve research questions (Kaplan & Duchon, 1988). The researcher creates categories and meanings from the data through an iterative process that starts by getting an initial understanding of the perspectives of that being studied. The understanding is then tested and modified through cycles of additional data collections. The objectives of this research were to examine the developing of a new semantic hub in a organisational environment, to see if there are factors that might give deeper understanding of the development of semantic hubs. The research method chosen for this thesis is an exploratory case study, because I want to explore the development process of a semantic hub.

3.2 Case Study

Benbasat, Goldstein and Mead (1987) defines a case study as a *examination of a phenomenon in it's natural setting, employing multiple methods of data collection to gather* *information from one or a few entities(people, groups, or organisations).* Case studies are often used to answer how and why questions as well as being the most desirable research method when the researcher has little control over the events in the case. It also is the preferred method to use when there is a contemporary focus with a real life context.

Where statistical research methods search for quantifiable data, the goal of the case study is to find new variables and questions for further research (Becker et al., 2012). Thus, the focus of my objective is not to find anything quantifiably different between the development process of semantic hubs and other information systems, but rather to find new interesting variables and questions, that might be beneficial to investigate further.

There are several types of case studies, and I have chosen to do an exploratory case study. The exploratory case studies are usually performed before implementing an investigation on a large scale. Their basic function is to aid in identifying questions and select types of measurement prior to the main investigation. The primary pitfall with exploratory case studies are that initial finds might seem convincing enough to be released prematurely as conclusions. This makes it important that every case study investigator works hard to report all evidence fairly (Becker et al., 2012; Yin, 2009).

Researchers may either conduct a single case study or multiple case study. Most case studies require multiple cases, but single cases can also be useful in specific settings, as for example when the case is extreme, unique, revolutionary, that it is a situation that has previously been inaccessible or it represents a critical case for testing a well formulated theory. A single case used for exploratory research can be followed by a multiple case study. Multiple research is most desirable since it allows for cross case analysis and generally yield more research material (Benbasat et al., 1987; Yin, 2009).

The case study for this thesis is a single case study on the development of the LogisticHub. In the beginning it was designed to look at one organisation within the single case, but due to changes in the research I had to expand to investigate the development of the semantic hub by including data sources from other organisations as well. In a way this made it a multiple case study within a single case.

One strength of the case study is the ability to use a large variety of data sources that might not be available in other research methods. The best way is to collect data from two or more sources so that they might be used to converge to support the research findings (Yin, 2009).

Documents - This is written material in a large variety, from newspapers to formal reports

Archival records - Can consist of organisational charts, service, personnel or financial records

Interviews - These can either be focused or open-ended interviews

Observations - Making observations in the field environment.

Physical artefacts - Can be a device, output tools, prototype.

The goal behind data collection is to obtain a rich set of data that embraces the specific research issue as well as it capturing the contextual complexities of the issue. The data collected for this research is further explained in Section 3.4

Case studies are needed to deal with creativity, innovation and context. A criticism of the case study research method is that it is difficult to make generalisations because of the inherent subjectivity, and because it is based on qualitative subjective data, generalisable only to a particular context (Yin, 2009). Yin (2009) states that the short answer to this is *that case studies, like experiments, are generalizable to theoretical propositions and not to populations or universes.* He also states that in order to ensure the quality of the research, the investigator must consider four points: construct validity, internal validity, external validity and reliability.

3.3 Why these sites

The same factors that dictates the use of single case also determine the site selections. Selecting sites should not be done by opportunity, but by the researcher. He must provide assurance that the organisations involved will not be harmed by its participation. An important part of this case study has been the organisation Statoil and their supply base at Ågotnes. The reason for choosing this site as a main focus in the thesis is that Statoil showed an initial interests in gaining and gathering new knowledge around the project LogisticHub. They are also one of the companies that have advanced the furthest in applying this new system. In addition, the supply base was closest at hand for consulting and meetings. Gathering information from other organisations have shown that almost all of them are aware of the new system that is coming. Most have a general understanding of what it is supposed to do, but a lot of the different stakeholders in different organisations are sitting on the fence waiting until everything is in place before taking a stand regarding the LogistichHub. A lacking understanding of the effects given by LogisticHub might also be a contributor for the little interest beforehand.

3.4 Data Collection

I have chosen to use several sources for data collecting. Thus increaing the chances for achieving validity and reliability from the result. I have chosen the following four types of data collections:

Documentation have been gathered from different sources, both publised and unpublished documents. A complete list of these documents can be found in sub section 3.4.3

Archival records have been given to me from one of the stakeholders, but since some parts of the domain covered does not exist in systems today, there were not that many archival records. A lot of these records have been added to a list as unpublised documents in section 3.4.3.

Interview/Meetings Meetings have been a large source of information for this study. Here different subjects have been discussed in a open minded matter. There was also held some interviews to gain a information in a more formal manner. Interviews are listed in section 3.4.2.2

Physical artefacts To test parts of the case study a simple prototype was tested, I looked at that simple prototype as a semantic hub artefact. The method for developing this prototype is explained in section 3.4.4

3.4.1 Interview as method

Part of the data collected from this master thesis comes through interviews and meetings with people working with the LogisticHub project. The collected data from these interviews have been analysed.

Kvale and Brinkmann claims that the aim of the qualitative research interviews are to capture the domain from the perspective of the subject, to find the meaning of the subjects experience and uncover their domain before capturing the scientific explanation. They further state that is important to try and uncover the perspective of the subject to truly describe the domain as it is viewed by the subject, with the understanding that what the subjects perceives is an important part of the true domain. The interviews and meeting in this thesis have been semi structured and theme based with subordinate question to fill in blanks along the way. As a researcher it is important to have an active listening role and to let the informant talk and describe their viewpoints and meanings as much as possible, only guiding their way through the interview (Silverman, 2006). Kvale and Brinkmann view the interview process as and knowledge production, where the interview in itself is an interaction between the interviewer and as such becomes a joint knowledge process. A challenge with the qualitative interview are that there is no sure way of conducting a good interview (Kvale & Brinkmann, 2009). There are no single solution as to how to gain the best possible result. Every interview situation is different and depends a lot on whom one is interviewing. Some informants are talkative, while others hardly say more than they need to answer questions. Research themes vary and question are perceived differently from informant to informant. The chemistry between the interviewer and his subject may vary from interview to interview.

3.4.2 Planning the case study

The case study first started with Statoil wanting to have a look at bringing another aspect to LogisticHub as well as increasing their knowledge on the subject, and additionally looking at how they might best utilise the semantic hub. Fist to look if inspection could be part of the LogisicHub, included in vocabulary and system to improve the existing inspection system that is very inefficient and paper heavy.

During the case study, the LogisticHub evolved and new issues became interesting and in wanting of evaluation. Security, how would one secure information through the LogisticHub? How could LogisticHub ensure that for example inspection that one organisation wants to share with another, would not be shared with all. This will be described further in the next chapter(4.2,4.3)

3.4.2.1 Planning interviews and meetings

The planning phase in this research came in several different stages. Since this is an ongoing project, and is always in development outside this research paper, the planning also had to develop. In the beginning it was important to gather as much knowledge as possible of the domain the research was to capture. This phase included gathering and collecting a lot of relevant literature and technical documentation concerning the domain.

3.4.2.2 Conducting interviews and meetings

Meetings were held at several different times and locations during the thesis. There were conducted 3 interviews with stakeholders from different organisations. One of these interviews where conducted in person and the two others were conducted by telephone. Telephone interviews are not ideal to have, but because of location and time constraints both from researcher and interview subjects it was decided to do the interview over the telephone. All three interviews were recorded and lasted for about 150 minutes. These recordings were transcribed and used for further analysis of this thesis.

During this thesis I have also had the aid of Nils Jacob Berland. He has been a main contributor and has acted primarily as a client connection and counsellor. He has many contacts in the offshore logistic industry, and has provided a lot of information gained through meetings and discussions on the thesis subject, that I could not be part of. Some of the meetings with Nils Jacob will be used as meeting with stakeholder, since they have often been based on information the was received in discussion the case with stakeholders in the oil and gas industry.

Meetings:

ME1 9. Mars 2012 with Statoil stakeholder. This was an initial meeting where I was presented with the case they wanted me to look at. The Statoil stakeholder is a member of the LogisicHub team where he is one of the contact persons for this project in Statoil.

ME2 23. August 2012 Statoil. This was the real start for the thesis, and we had a meeting to go over the case again and discussed the viewpoints they had. As well as I got a tour of the facility and was explained the process of inspecting the Z-015 equipment.

ME3 27. September 2012 Meeting with Nils Jacob. Where he had been discussing the use of Semantic Web technologies, versus using some existing XML tagging system, as a solution for the semantic hub. As well gave some feed back on the system narrative I had written.

ME4 15. November 2012. Meeting with Nils Jacob, where I was informed that there was a new document released.

ME5 18. December 2012. Meeting with Nils Jacob. Where he brought back information from several different stakeholders. On how they have responded to the release of this document, and what their thought on the next step in the process .

ME6 9. January 2013 Stakeholder Statoil.

ME7 25. January 2013. Meeting with some developers developing a simple semantic application that is suppose to utilise the LogisticHub web service to visualise the event data.

ME8 12. February 2013 Group meeting at Statoil, where the the semantic application where presented in front of several stakeholders.

ME9 20 April 2013. Meeting with Nils Jacob, discussing resent updates in the LogisticHub development process, and getting feedback on the viewpoints of CCU owners and handlers of the LogisticHub, along with the their thoughts and ideas.

Interview:

IN1 Mars 2013 Statoil and LogisticHub stakeholder

IN2 April 2013 ConocoPhillips stakeholder

IN3 April 2013 Leading stakeholder LogisticHub.

3.4.3 Documentation & Archival records

Through the process of this case study and in the planning of this case study there has been a review of relevant documentation. Some of the documentation is relevant and some documents less relevant. A lot of these documents are standards made to handle different aspects of the logistic handling in the offshore supply chain. There were also some drafts of different diagrams showing purposed solutions to certain task handling. Because the project LogisticHub is an ongoing process, some of the documentation has changed and more documentation has appearer during the research. This has affected the outcome and focus. Specially, a document was released, describing the initial specification and requirements for the LogisticHub (The Norwegian Oil and Gas Association, 2012).

Different archive documents were reviewed. Most of them were data examples of the tasks and processes from the existing systems. Since most of the LogisticHub processes does not exist to day, these archive documents mostly related to inspection and the storing of inspections.

These are the different documentations and archive document used in this thesis.

3.4.3.1 Published documentation

PD1 'NORSOK STANDARD Z015: Temporary equipment', the standard document for the Z-015 temporary equipment written and managed by Norsk Standard (2004).

PD2 *OLF GUIDELINE No.112 DEPLOYMENT OF RADIO FREQUENCY IDENTIFICATION (RFID) IN THE OIL AND GAS INDUSTRY*, A complete set of documents discussing many aspects of deploying RFID technology into the offshore supply chain, and the basis for the LogisticHub. These documents were the basis for creating the LogisicHub. They provided me with much insight in how the RFID track and trace system was thought to be developed. This is a very large document consisting of 9 parts.

PD3 'Logistichub Requirements Specification', requirements and specification published in November 2012. This document became a turning point in my research. At the point of this release the case made a changing turn. Stakeholders got some of the answers they were looking for and others appeared. The document also helped me validate some of my early findings. By giving me something that I could compare findings to, showing that my development process was on a similar path as the LogisticHubs development process.

3.4.3.2 Unpublished documentation and archive documents

UD1 Document from Statoil, is the start of requirements and specification document, for a new system. Built to handle business process surrounding the Z-015 temporary equipment. This system was suppose to be open and shared between different organisations. This initial document served as a kind of semantic hub specification for the initial parts of this research. I handled this document as a customer request for a semantic hub.

UD2 Video recording of a user talking me through and describing the existing Lotus Notes system for handling Z-015 temporary equipment. I felt that a data dump of the database would not give me much information that other documents did not have. So I asked to get a preview of the system and was provided with this Video explaining the existing system. This gave me a unique insight with commentary.

UD3 Filled out checklist from the existing system A. This shows a printed checklist from the existing system. The checklist form is given as a fixed form in the official Z-015 document (PD1), but I wanted to see how it looked when it was filled out.

UD4 Presentation given on LogisticHub by Dr. Thore Langeland 22.11.2012[1]. This and and UD5 are both presentation about the LogisticHub. These documents gave me some deeper insight into the progress of the project as well as it giving me insight into the people planning the LogisticHubs desired outcome.

UD5 Presentation given on LogisticHub by Dr. Thore Langeland, send to me by mail from Thore Langeland.

[1]https://www.posccaesar.org/svn/pub/LogisticHub/121104_LogisticHubRequirementSpecification.pdf

3.4.3.3 Email correspondence

During the research, question that I felt needed answering arose. Due to geographical and time restraints, I forwarded these to the stakeholder via email. I was also given insights into other peoples questions and discussions regarding LogisticHub. This was useful useful for the research. Not all the emails are listed in this sections, because some of these emails do not contain relevant information or because they contain confidential information.

MA1 Mail correspondence with Statoil stakeholder. Answering questions on the process of handling Z-015. As getting the video recording.

MA2 Mail correspondence with ConocoPhillips stakeholder. Answering questions about the how they handle Z-015 in their existing systems.

MA2 Mail containing information about the ontology building.

3.4.4 The Simple Prototype

As part of the investigation I chose to develop a simple prototype. This by starting to use the development techniques to outline the prototype. As part of the investigation and designing stages of the development process. This was used to develop a simple prototype of a semantic hub, implementing different these different aspects of the case study. I felt that part of the research objective could best be investigated through building and testing a prototype, in addition to delays on the LogisticHub development process that forced me to start a development process on my own.

Chapter 4

LogisticHub, a Semantic Hub

By examining the development project of the LogisticHub, this thesis aims to investigate the process of developing this semantic hub in the Norwegian offshore supply chain. I was first presented with a case by Statoil. They wanted to get a deeper understanding of how a semantic hub works, and presented me with a case to examine. However, during the research, a requirements and specifications document was released, and the interest and focus changed in the research changed.

This chapter starts with a short review of the LogisticHub and the thoughts behind it. Next it will describe the initial case presented to me by Statoil, then the effects and changes in the case after the release of the requirements and specifications document. Then I will describe sharing and ownership of data.

4.1 The LogistichHub

For a long time there has been work on elucidating the use of of Radio Frequency Identification (RFID) technology on equipment and cargo in the Norwegian oil and gas industry. This to create an easier platform to communicate digitally between business events and the more tangible handling of these events, such as moving and inspecting of equipment and cargo. Since the handling of this is aimed at supporting the whole of the Norwegian oil and gas industry, there should be sharing of events across the different organisations. This because many of the events like moving cargo is handled by different organisations and systems.

As a result of this, EPIM proposed to create a semantic hub called LogisticHub (The Norwegian Oil and Gas Association, 2012, PD3). The main goal of the semantic hub is to share the equipment and cargo data collected from the RFID technology between the different organisations on the Norwegian continental shelf. This so the organisations can obtain more real time high quality information. The LogisticHub will in the first phase be developed to manage and capture tracking information from cargo carrying units (CCU), by storing events that are triggered by the CCUs as they move through the life cycle of the offshore supply chain.

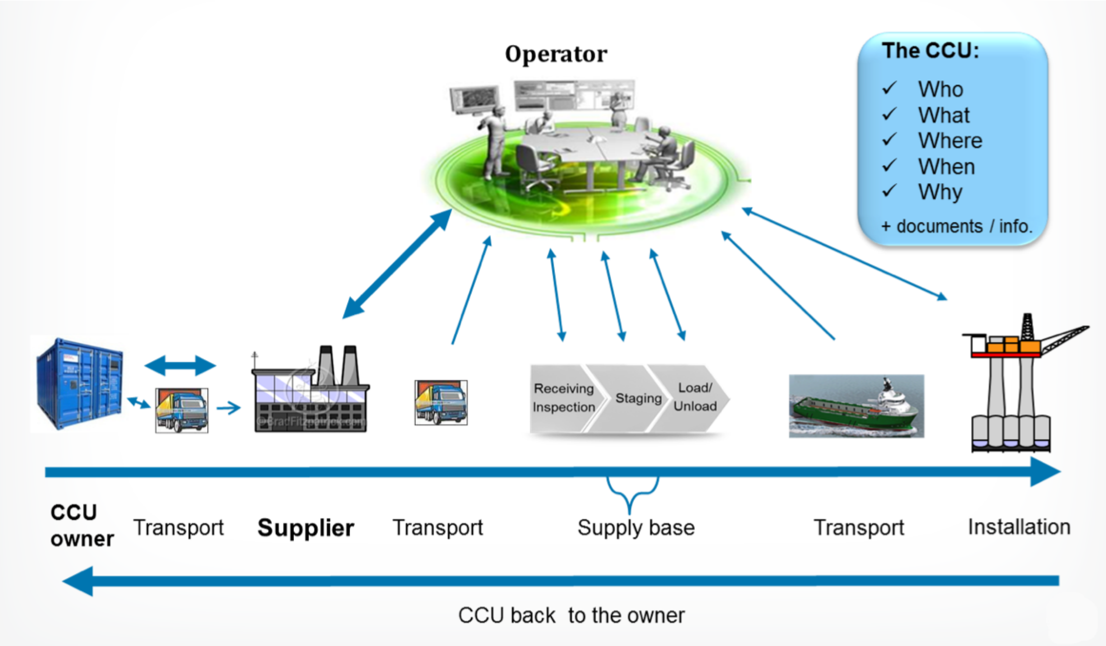

FIGURE 4.1: The life cycle of a CCU in the offshore supply chain (The Norwegian Oil and Gas Association, 2012)

LogisticHub will provide its users with this data in two different ways. One way will be to provide the users with a web service. This web service will be able to receive requests to, push data, or get data from the semantic hub. The second way will be to provide the users with a web interface, which the users can use to query for the data they desire. In the first phase the queries will only be available as predefined queries or created from parameters. The main advantage in proving a web service to the users is that this organisations to embed the web service into their own existing system, and thus can use the data collected in handling with internal operation (The Norwegian Oil and Gas Association, 2012, PD3).

Most data in the LogsiticHub is new and created from triggered events by the RFID technology. The other data stored in the semantic hub will be master data. Much of this data will scraped and converted into RDF from several different data sources by the LogisticHub. The data that is not openly available to the LogisticHub, will be provided by the different organisations. Organisations will maintain and update the data they are responsible for through the web service. The web service will receive and respond to the requests in XML.

Thus making the semantic hub the centre of information, where business partners connect to share information. This type of system for handling tracking and tracing of cargo and equipment, does not exist as of today. The LogisticHub is an ongoing development project. During the duration of this thesis, the LogisticHub evolved, and as a researcher I had to adapt to these changes as the development evolved.

4.2 Initial view on LogisticHub

The initial view that was presented to me by Statoil, was that they wanted me to examine how a semantic hub works, and how the semantic hub could best be utilised by their organisation. To examine this they wanted me to look at the Z-015 temporary equipment, and the relating business processes of handling the Z-015 in the offshore supply chain. They found the handling of Z-015 lacking in existing systems, so at the same time, they also wanted to investigate the possibilities of including these business processes into the semantic hub. The process they stated as the most inefficient and unproductive, was the required inspections of the Z-015. This due to lack of interoperability with this process both internally, within their own organisation, and externally to externally to other organisations. They wanted me to look at creating an ontology for the Z-015 equipment, as well as look at the business processes surrounding the Z-015 equipment.

4.2.1 Z-015 Temporary Equipment

Temporary equipment used on the Norwegian continental shelf is defined by a NORSOK standard Z-015. It defines the minimum technical and safety related requirements for temporary equipment used on installations (Norsk Standard, 2004). Norsk Standard (2004) defines temporary equipment as equipment used for a limited period of time, and that requires connection to an offshore installation before use. Therefore, this type of equipment has to go through a thorough check before it can be approved and shipped offshore.

4.3 After Requirements Release

Early November, a requirements and specifications document were released on the LogisticHub (The Norwegian Oil and Gas Association, 2012, PD3). This document included information about some of the things I had looked into. The release of this document also changed the view presented by the stakeholders. One of the biggest stakeholder changes of focus where towards security of data within in the semantic hub. How could their information be securely stored in the LogisticHub. So I started to look into several different access control handling frameworks from Semantic Web technologies.

A CCU moving through the the supply chain will create several events. These events will be pushed to the LogisticHub. Each event consists of the following: who, what, where when why? Many of the things in the document reflected things I had already thought of and work out. One thing I had already covered was describing what organisations triggered event. As for what type of event that is triggered, I had partially covered this in the existing ontology and UML class diagram model. Order became and hiring event and inspection became a document in an inspecting event. `Inspecting` and `hiring` became subclasses in a super class `event`. I made a decision to not model all the other events described in the The Norwegian Oil and Gas Association (2012, PD3) into the object model as it would cloud it up and not show anything new or interesting, but these events will be written into the ontology.

The main reason for adding the class of documents was that at a meeting (ME5) with Nils Jacob , he informed me that different stakeholders in the offshore supply chain had expressed an interested to be able to link to important documents in the events, like inspections, hook-up schemas, certification papers. Taking this into consideration, I looked up the documents that I were not previously familiar with and found that these documents needed to be modelled into the ontology (Norsk Standard, 2013).

4.4 Sharing & Ownership

One of the main purposes of a semantic hub is to provide a connective link of information between organisations, by providing data from several heterogeneous sources. This data should also be provided dynamically, to give its users real time information. An ontology enables the possibility of sharing data across organisations and systems, without the data losing its meaning. The data can either be stored directly in the semantic hub or be linked to from it. This will enable an organisation to gain access to data shared by other organisations, but not all data that is stored in the semantic hub is desirable for all to have access to. It has been expressed in several interviews(ME1,ME2,ME3), that organisations need to be sure only that information that they want to share with one business partner, is not accessible to others as well. This Makes it important to clearly define the ownership of the data stored in the semantic hub, and who has access to what data.

ConocoPhillips and Statoil stakeholder(IN1,IN2) both stated that they are not completely sure what data they desire stored in the LogisticHub, and that they need to make deeper investigation before they can be sure. The data stored in the LogisticHub is organised in two groups, master data and event data. Master data is the core data in LogisticHub that ensures the that the business operations can be performed. The data registered about an organisation, for example its name or address, can be seen as that organisations master data. This master data is open for all and is not restricted by access rights. Therefore this information should not contain anything sensitive, and as stated earlier will be mostly collected from openly available sources. Event data is the data generated by the RFID technology. Event data represents the business events triggered by or for an organisation when handling a CCU in the offshore supply chain. This data is stored in the LogisticHub, but is not openly accessible by all organisations.

Owners of information and data is defined in the 'Logistichub Requirements Specification', PD3 as *organizations which produces information regarding the identifiable item.* The owner, hirer or custodian organisation that created the event, will be registered as the information's or data's owner, but other organisations might have access rights to the information. These are also stated in the requirements for LogisticHub (The Norwegian Oil and Gas Association, 2012, PD3) and are listed as four rules:

- CCU owner has access to all events generated by her/his CCUs where there are changes of custody (loading/unloading) and events of the CCU on the way to its custody (loading and departing previous site)
- The hirer has access to all events generated by her/his hired CCUs
- The operator has access to all events generated by CCUs in her/his custody and events of the CCU on the way to her/his custody (loading and departing previous site)
- The supply base owner has access to all events generated by CCUs in her/his custody and events of the CCU on the way to her/his custody (loading and departing previous site)

A common denominator in these rules are the events. The different organisations have rights to access different events and thus the data stored in these events. In the ontology, I have modelled the inspection as a document in the event inspecting (6). When two organisations are asked about sharing inspection data, they both seem interested. However, one states that he needs to take a closer look to see if there is any real value in it, while the other seems to think their could be great value. Furthermore, one stated that I would have to look at the ownership of the data in order to see if this sharing would be possible at all. Thus identifying the ownership of the data in the semantic hub is essential for there to be interoperability in the semantic hub. This is especially if the semantic hub is supposed to support sharing of information outside the four rules. As example with the inspection document.

Chapter 5

Development techniques used with the Semantic Web

This chapter will examine the use of different system development techniques in the frame of the initial case presented. This as a step to figure out which development techniques could fit with the development process of a semantic hub. Since development techniques are considered valuable for the development process, exploring the use of some development techniques in development of a semantic hub, should provide a deeper understanding of the whole development process (Avison & Fitzgerald, 2006).

I gathered information from meetings and documentation (3.4.2.2,3.4.3.1,3.4.3.2). This information was applied to investigate and analyse the semantic hub through the use various development techniques. Initially the goal of the research was to closely follow the development of the LogisticHub, but due to delays in the project and the time restraints in the master thesis, I had to start off exploring parts of the development process by my self, starting with the system development techniques. Why these specific development techniques were chosen will be further explained in the corresponding sections.

This chapter starts by looking at how building of an ontology, with the help of UML. Then it examines how business processes handled by the semantic hub can be modelled with the help of BMPN.

Even though the system development methods here are described in sequence the different system development methodologies were developed parallel in several iterations. Thus the results from one system development techniques affected the other.

5.1 The Road to Ontology

The overall desired outcome of the LogisticHub is to provide interoperability between different organisational systems in the offshore supply chain. This can best be achieved by providing syntactic, structural and semantic interoperability between the systems. This can be achieved by building a ontology. When building ontologies it is recommended to follow certain development methods with several steps. Some of these steps can be seen as a development techniques.

Initially, it is recommended to gather as much information as possible about the domain in which the ontology will be based. In a meeting (ME1, ME2), Statoil presented the case of the Z-015 temporary equipment, and the handling of this type of equipment in the offshore supply chain. Several different documents and archive information were reviewed (Norsk Standard, 2004; Norwegian Oil and Gas Association, 2010, (PD1,PD2)), as well meetings and interviews with stakeholders interested in the LogisticHub. From these sources, I created some competency questions to narrow down the scope and domain of the ontology:

What type of unit is it? Stakeholders in Statoil wanted to be able to scan a unit, query it to get its type, to get the correct inspections for that unit (ME1,ME2,UD1).

What is the units id? The unit needs to be identified, and the stakeholders wanted to have the unit identified, both with their local id, and id as a URI (ME2,ME3).

Where is the unit? The main goal for creating the LogisicHub is to have track and trace on units. Thusm it is important to know where the unit is located (ME1,ME2,IN1,PD3).

What checklist for this unit? From type the stakeholders in Statoil wanted the possibility find the correct checklist for a specific type of unit (ME1,ME2,UD1).

Does the unit fit with requirements of the order? During a meeting, a stakeholder in Statoil expressed a desire to get more automation in processes related to handling of equipment of the type Z-015. Thus this might help the owners find Z-015, to fit possible orders, or help hirers better find matches for renting (ME1,ME2,PD1).

Is unit ready for shipping? Stakeholders expressed through interviews that they desired a reduction in time delays in the shipping process. They wanted to be able to get information from the system, for example when units where ready to be shipped (ME1,ME2,ME2).

Who owns the unit? To know were the unit is from and thus know to whom messages should be sent, if that would be needed. The owner of a CCU also have rights to certain event data generated by their CCUs (PD1,PD3)

Who ordered the unit? The person that ordered the unit should get a message if there is a problem relating to the delivery of the unit (ME2,PD1).

Where is the unit going? Many organisations have several places where units are shipped to and from. It could therefore be useful to know to what location it is going (ME3)

What grade did unit X get in the inspection? The stakeholders want relevant personnel to get messages if the unit gets a certain grade in a inspection (ME2,ME8).

Who inspected unit X? All inspection documents need to be signed by authorised personnel (ME2,PD1,UD1).

When was the inspection done? It is necessary to know when an inspection was done, to be able to trace back to it

What zone is the unit approved for? For safety reasons, units need to be approved for certain zones of use (PD1,UD2,UD3).

When does the certificate go out? In a meeting there was stated an interest for being able to check the units certificates, and when these might expire, or when they where issued (ME1,ME2).

What Hook-up does this installation use? Offshore facilities have different ways to hook-up units to power, communication etc. The hook-ups for each installation are described with a schema on a web site. Stakeholder stated that these hook-up schemas could be useful to have connected to the semantic hub (Norsk Standard, 2013, ME2,PD1).

How do I hook-up this unit? Similar to the point above, the stakeholders desired documentation-like guides to hook-up units to be digital (ME2)

Are all the documentation here available? The inspection requires a set of documentation to follow the unit. There was also expressed desires to have the documentation following a unit digitally linked to it (ME1,ME2,ME9).

What maintenance have been done on the unit? Offshore facilities need to have maintenance logs on the units. There was expressed that accessablility to these logs where not well covered today and these could be useful to have access to on land(ME2,UD2).

5.1.1 System Context

As stated by Noy and Mcguinness (2000), ontology building is an iterative process and will be changed throughout the entire development process. To better understand the processes in the existing systems, I chose to first write a system context overview in the form of a narrative. I gathered data from how the existing system works today and wrote the story, guiding one through the different tasks, users and actions describing these in detail. When this was done I changed the story by taking the information gathered from meetings and documentation (3.4.3,3.4.2.2), and tried to fulfil the desired improvements given by the stakeholders. I used this system context as the main source for which to gather the key terms and concepts. When these terms where gathered, the context description was revised and modified to fit feedback given by stakeholders (ME3). Afterwards the terms where revised and modified according to the new context written.

When developing the ontology and finding the terms and concepts, I chose to follow a combination of techniques. I mapped out the terms and concepts in a middle out perspective, and with object modelling. I also used the competency questions above as assurance that the correct terms and concepts were captured. I ended up with a initial list of terms and concepts. These terms and concepts will be explained in more detail in the next section (5.1.2). The list of terms was reviewed and changed iteratively, using the other system development techniques and through communication with stakeholders to determine changes.

- Unit
- Inspection
- Checklist
- Control item
- Supplier
- Inspector
- Installation
- Requisitioner
- Order
- On board log
- Goverment inspector
- Z-015 temporary equipment
- Offshore installation manager

5.1.2 UML Class diagram

After having written down a set of basic terms for the ontology, I chose to model these terms into a UML class diagram. This because UML has a standard mechanism for defining extensions for specific application contexts similar to ontology modelling. UML is also widely adopted in the system development industry, and is taught at several universities. My reason for using UML class diagram, was to find a way to more easily capture the relationship between the different terms. This would also provide a visual representation of the terms to work with. This visualisation can also be used to more easily communicated to the stakeholders (Kogut et al., 2002).

The UML modelling follows the recommendations of Fowler (2004). The UML class diagram was first developed prior to the release of the LogisticHub requirements documents (The Norwegian Oil and Gas Association, 2012, PD3), but has been modified after its release to better fit with new interests and requirements. The initial class diagram is shown in a draft in Figure 5.1.

The model was extended with new terms and relations that emerged after reviewing the new document, and by reviewing the model with the competency questions. These terms and relations were put into a new model. Figure 5.2 shows the final model diagram that is the basis for the ontology.

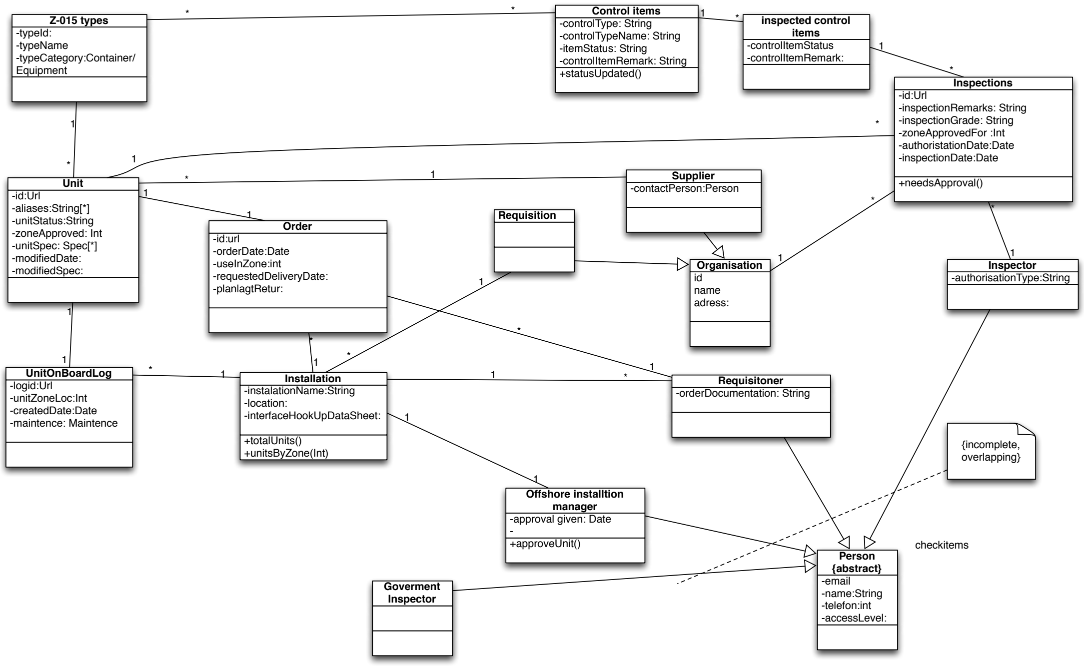

FIGURE 5.1: A first draft of class hierarchy diagram drawn from system context

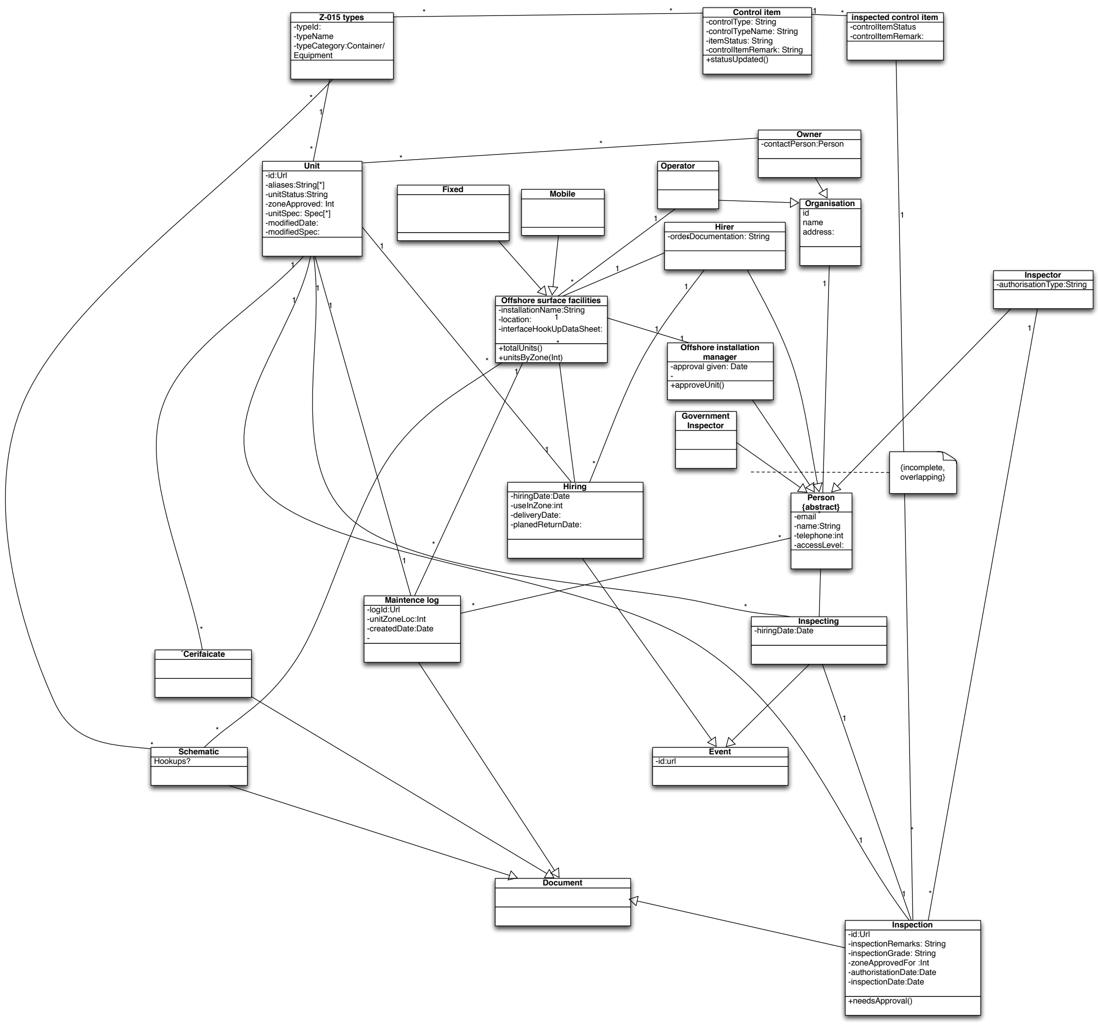

FIGURE 5.2: The final class diagram, which was used to code the final ontology

The terms displayed in the final draft why they were chosen, and what changes that were made from the draft, to the final diagram are described in the list below.

Unit is here the tangible thing. A unit does not need to have a Z-015 type, but if it does it can only have one. Initially the thought was that only one organisation would own a unit, but with the release of 'Logistichub Requirements Specification', (PD3) owner relationship became more elaborate. It states that a CCU owner may hire a CCU from another CCU owner before hiring it out to a supplier or operator, and in this case they are both owner of the CCU. Thus a unit can in some special cases have two owners. A hiring can only have one unit at the time, and a unit can only be part of one inspecting event (Norsk Standard, 2004, PD1). It was expressed at a meeting with Statoil stakeholder (ME3), that it could be useful to have access to the certificates, and other useful documents, belonging to the unit from the semantic hub. These were added as document in the model, where for example a certificate can only be connected to one unit, but a unit can have as many certificates as it needs. Documentation (Norsk Standard, 2004, PD1) shows that one of the documents is a maintenance log, for the events of a unit from when it arrives at the offshore installation, to when it departs. In initial meetings with the Statoil stakeholder there was talk about lacking coverage of these maintenance logs in existing systems. I chose therefore to also implement the maintenance log document into the knowledge base.

Z-015 Types is the different classification given to a type of temporary equipment or container defined in 'NORSOK STANDARD Z015: Temporary equipment', (PD1). This class has several subclasses that are not modelled in the diagram because they do not enrich the meaning of the diagram. Each of these Z-015 classifications has several control items that combined become the checklist form. This checklist is filled out by the inspector as he performs the physical inspection. These control items can be found in a matrix displayed in the 'NORSOK STANDARD Z015: Temporary equipment', (PD1). The different types of Z-015 are required to have documents that need to be available with the unit. Today the documents need to be with the unit physically in paper form. During a meeting with Statoil stakeholder (ME3) is was expressed that these documents should also be available digitally and linked to unit or Z-015 type.

Control item is one item in the checklist that needs to be checked and controlled for the unit to finally get rated. Each type of Z-015 has a different set of control items that need to be checked. In the 'NORSOK STANDARD Z015: Temporary equipment', (PD1) there is a matrix between the different control items and the type of Z-015 unit.

Inspected control item is the control item stored and checked. It contains the information about the item's status and a remark if a control item needs commenting by the inspector.

Inspection is the document that is linked to from the inspecting event. An inspection document consists of a list of completed control item, the inspector that inspected the unit, the inspecting event it was inspected under, and the unit inspected. An example of how this inspection document looks can be seen in appendix A. In the initial draft, the inspection document was called checklist. This was changed in the final diagram due to the ambiguous meaning of the word. When discussing the checklist at a meeting (ME2), it became clear that the checklist only referenced to the control items and not the other information that is required to be in the inspection document, by the standard given in 'NORSOK STANDARD Z015: Temporary equipment', (PD1).

Inspecting is the event or task of conducting an inspection, and a subclass of event. In the event, the inspector fills out the correct form and sends the filled out form to the LogisticHub. In the first draft of the diagram, inspection was thought of as the event of inspecting, but with the release of the LogisticHub requirements they introduced events as business processes, and I had to model this into my class diagram, and the name inspection became the name of the document.

Hiring is the event that represents the start and end of the rental of a unit. The hiring starts with a hirer sending a request for a unit. The unit owner and the hirer agrees on a rental and the hiring events starts. In the first draft and in the Z-015 documentation (Norsk Standard, 2004, PD1) this was called ordering. I changed this to make it fit with the LogisticHub business process events. This is also a subevent of events.

Hirer is the person or user that requests the hiring of a unit. Similar to hiring this has a different name in the initial draft, but was changed to fit with the LogisticHub terminology. Previously called requisitioner.

Owner is the organisation that owns the unit. The owner can own many units. The owner used to be called the supplier, but trying to follow the terminology of the LogisticHub (The Norwegian Oil and Gas Association, 2012), I changed it to owner. The named was also changed due to the fact that supplier sometimes refers to the supply bases.

Offshore surface installation is the location where the unit is shipped to. The hirer hires the unit for an installation, making that installation the delivery point for the unit. An installation may have many units on hire at the same time. The maintenance logs of the units are connected to the installation. I a discussion at a meeting, it has been pointed out that an installation might in some cases be either mobile or fixed in their position (ME7). I chose to model as subclasses of offshore surface installation, because it might be interesting to know were geographically the mobile unit is located, since this is not a fixed position. This in order to best determine supply base for delivery of the units. All installation have an offshore installation manager (OIM). The OIM is the head of the installation and is the decision maker. If the inspection of a unit gets a too low rating and need extra approval, the OIM is the one that needs to approve the unit.

Offshore installation manager (OIM) : is the chief on the offshore installation. He needs to get a message if a unit does not meet the criteria for the order, so he can approve the units faults or reject it. Often he might also be the hirer of the units, but not always. Thus a person can have several roles in the system.

Inspector is the person or user that performs the inspecting of the unit. The inspector often comes from a third party organisation, and not from the organisation the inspecting is done for. The inspector needs to be authorised in order to be able to perform the inspection, and will need to sign the inspections.

Operator is the organisation that owns the offshore installation were the units are used.

Event is triggered by an organisation, and can often directly be connected to a business process. Event is a superclass of inspection and hiring. From the LogisticHub requirements document there were several other events stated. These are not modelled into the class diagram, but are going to be written into the ontology.

Organisation is a superclass of the different organisations, like operator and owner. This class is the superclass of operator and owner, but there could have been several other subclasses in organisations like inspections organisation, transporter organisation. I have chosen to limit organisation subclasses to only those that I have explained here.

Terms and concepts become the basis for the ontology, that is the foundation for distributing the meaning of data between different systems and applications. To better grasp the semantic hub system, I needed to explore how the information is distributed between the different systems. In a system that is supposed to provide interoperability, one needs to explore who is communicating to who, as much as what is communicated. To examine this I chose to use the Business Process Model Notion (BPMN) to help identify and capture the communication flow.

5.2 Business processes to flow

The essence of modelling business processes into task is to get a clearer picture of the system, to ensure that the right user executes the correct tasks at the right time. In complex business systems there might be several different applications working together to solve one business process. In this case the semantic hub will be used by several systems to complete tasks. By modelling the tasks and information flow between the different tasks, one will get a clearer picture of what needs to be communicated. These models can then be used to further validate the ontology in a similar fashion as the competency questions (Object Management Group, 2011).

The business process modelling was mainly developed before the release of the LogisticHub requirements and specifications document. The release of this document made impact on the model, as will be described in the sections below.Roles were collected from the object model and system context narrative. Tasks were also found in this context narrative as well from other sources.

The BPMN will break down the different aspects of the business process in the handling of Z-015 units. This is one of the more complex processes that is not completely covered in the LogisticHub requirements (The Norwegian Oil and Gas Association, 2012, PD3), and is stated by several important stakeholders from both operators and owners, as a business process that has a great need for improvement (ME2,ME3,ME9). The flowcharts were created by collecting information from documentation, archive data, meetings and observations in the field. The results of these charts are an analysis of the data collected from these different data sources.

5.2.1 Hiring & Check

The start point for my flow chart was a simple flow chart displayed in the 'NORSOK STANDARD Z015: Temporary equipment', (PD1). This flow starts with a hirer making an order and sending it to their contract department. The contract department then haggles out a deal with an unit owner. When owner and contract department agrees, they draw up a contract. This is when the hiring event starts.

Feedback from meetings told me that this process was already covered (ME5) in existing systems and that the operator did not want this process in the semantic hub. Therefore in the next BPMN chart the event starts when the order has been approved, and when the hiring event is supposed to be added to the semantic hub.

The final BPMN draft of the hire and check, shows the interaction and connection between the different actors in the semantic hub, and their systems. This model has three pools with two lanes locations and a fourth pool that is the semantic hub in this case the LogisticHub.

The first pool, called `owners` represents the organisation that owns the unit. Owners are on a different system than the two other lanes. This pool has two lanes, `office` and `inspector`. Office represents all those tasks that are done at the owners office, such as paperwork and appointments. The inspector lane represents the person that performs the inspection. Inspector is modelled into a separate role because the inspector comes mostly from a third party organisation, while the inspection is done at the owners (IN2,ME2). Inspectors could also have been drawn as a role in the LogisticHub pool, and thus the task of inspections are part of the LogisticHub system. This might have been some advantage in obtaining better interoperability for the different organisations, and especially between different inspectors, since the inspection system could be the same for all when it is handled on the semantic hub. As long as the forms the inspector fills are modelled after the Standards of Z-015 (Norsk Standard, 2004, (PD1)), what system they use should not matter.

Supply bases can be owned and operated by different organisations. Some bases handle the offshore supplying for many organisations. Large operators own their own supply bases. Thus the pool of base `base/operator` might work on the same system or different from the pool of `offshore installation`.

The base/operator pool has two lanes, inspector and base personnel. Base personnel will handle the unit on the base and manage it when in their custody. The inspector will do the same as above in the owners pool(Norsk Standard, 2004, ME2,PD1).

Offshore installation is the location in which the unit was ordered. Some offshore installations can be rented by operators, and these rented installation uses systems that are not compatible with the operators enterprise system, making it a possible for an offshore installation to be outside the operators systems (ME2). The pool has two lanes, the hirer that requested the unit and the OIM that is the chief of the installation. In some cases the OIM and the hirer might be the same user, but since this is not always true they are drawn as two lanes (Norsk Standard, 2004)

The flowchart starts with the hiring event being agreed upon between the owner and the operator. 'Logistichub Requirements Specification', (PD3) states that the *"hiring event is a CCU owner contracting a CCU for hire and providing the information to the LogisticHub"*. Therefore the model starts with the hiring information being added to the semantic hub by the owner. When the owner creates the hiring event, there is a check done to see if the unit is already registered. If the unit is not registered it needs to be added to the semantic hub, in this case the LogisticHub. In an unpublished document (UD1) from Statoil, they show that the base is responsible for adding the CCU information into the semantic hub. I chose to follow the 'Logistichub Requirements Specification', (PD3), that states that the owners are responsible for adding information about their own units. Therefore when the owner attempts to register the unit hiring, the system checks against the semantic hub. If it is already registered, the process moves on, if not they are requested to add the CCU information. In the case of Z-015 units, a unit can be modified or rebuilt to fit the hirer's requirements. Therefore it is possible that if the unit is previously registered it might need its data changed in the semantic hub. This would only be relevant if the Z-015 changed classification. This might put demand on the system to also check for rebuilding and modifying of certain units. When the unit is registered the owner will then issue an inspection. By checking if the unit is registered in the same process as ordering, will ensure that inspectors do not have delays when performing the inspection or in worst case make a wasted journey to inspect units that are not registered (ME1,ME2). The inspection task represents the sub-process of inspection, this is explained in more detail in the next section, and can be seen in Figure 5.4. If the inspection fails, the owner must start a new hiring event with a new unit. When the unit passes inspection the unit gets updated to ready for shipping, and the inspection data is sent to the LogisticHub. Next, the unit departs from the owner, the departing event is triggered and the departing event data gets sent to the LogistcHub.

The next activity is at the supply base, where the base needs to find an inspector, for the equipment and schedule an inspection date. When the unit arrives at the supply base, the arriving event is triggered and stored in the LogsicHub. The inspector then gets informed of the units arrival, and inspection can be performed as described in Section 5.4.

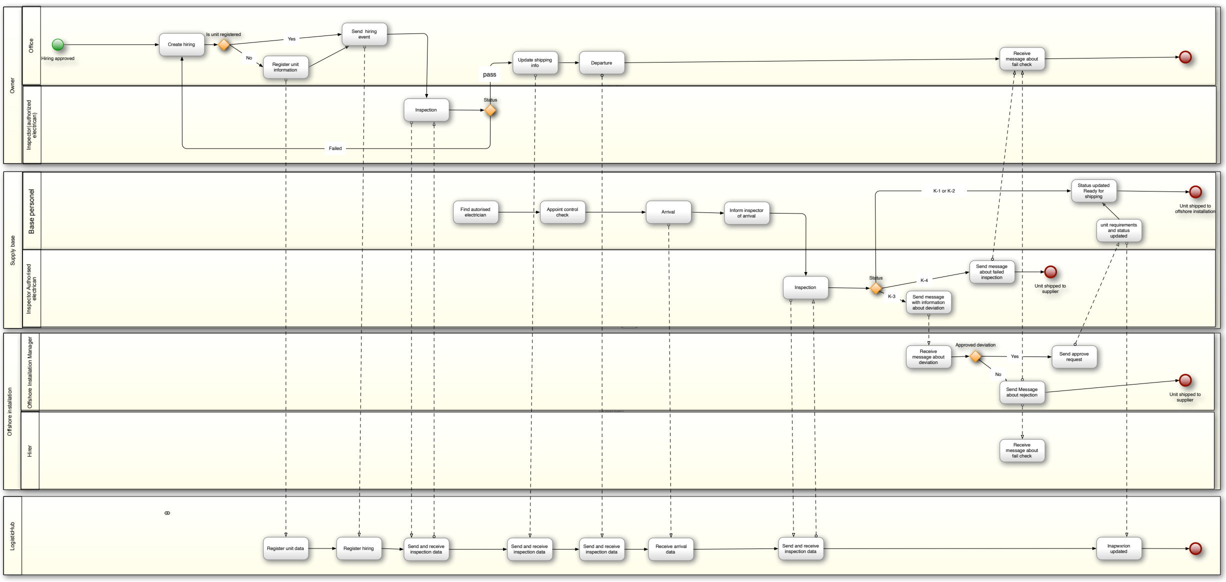

FIGURE 5.3: A first draft of class hierarchy diagram drawn from system context

When the Z-015 inspection data is sent to the LogisticHub it can be graded into four different grades. A perfect grade is K-1. This means that the unit has no faults and can be shipped offshore. K-2 is also a passing grade, but with K-2 the unit has some faults, but not serious enough to become a safety risk, and can thus be shipped offshore. When the inspection gets one of these two grades the unit gets automatically updated to ready for shipping, and the flow ends with the unit shipped or departed for installation. This triggers a new departing event that is sent to the LogisticHub. If the rating is K-3, the unit did not pass the inspection, and needs extra approval before it can be shipped. Therefore when the grade K-3 is given, a message gets automatically sent to the offshore installation manager, that this unit got a rating of K-3 and needs extra approval before it can be shipped. If the OIM approves the unit, and the inspection gets updated with the OIMs approval, it then gets shipped offshore. If the OIM rejects the unit, the unit gets shipped back to owner, and the hirer get sent a message that the unit did not pass the inspection and was rejected. K-4 grades the unit with serious faults and the unit is automatically rejected. The hirer and owner get messages about the failed inspection and the unit gets shipped back to owners location. During these processes the LogisticHub gets updated with the inspection document and the different departing events.

Stakeholder stated in several interviews that inspections are an important focus. The inspections are especially intricate since they are mostly done by third party organisations, and the inspection require data from both owner and hirer to be complete for a unit. Inspectors mostly use paper when doing an inspection and then the information gets entered into a database. Therefore, I decided to create a sub-process of inspection, and thus get a closer look at the communication and handling of the tasks in it.

5.2.2 Inspection flow

Inspections of Z-015 are complex because a complete inspection can be seen as two stages. These two stages are done by different organisations at different sites and systems. In the first stage the Z-015 is inspected by the Z-015 owner, to make sure the unit is up to the safety standards set before it can be sent to the hirer. The second stage of the inspection is an extra safety measure, where a new inspector inspects the Z-015 at the hirer location, to ensure that the previous control was correct and that the unit still is within the safety regulations, before it can be shipped offshore. The inspection in stage two needs the data from the inspection in stage one.

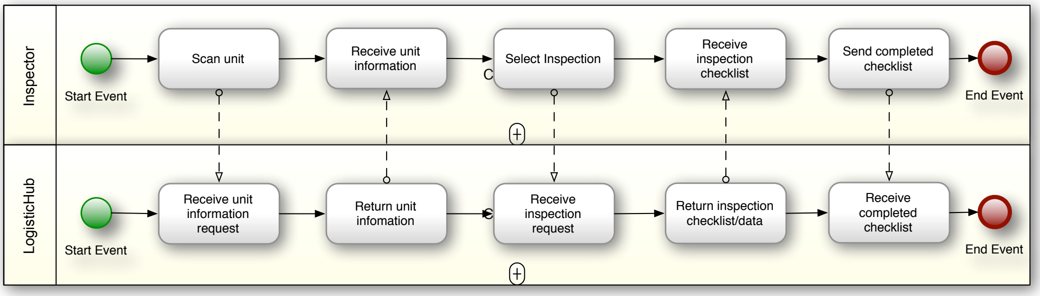

FIGURE 5.4: Flow chart of inspection

The flowchart seen in Figure 5.4, shows the process of inspecting and the flow of information between the LogisticHub and the inspector. The BPMN model has the lane inspector, that is the user that performed the inspection, and LogisticHub that is the semantic hub. The model starts with the inspector just having logged into a system. The inspecting starts scanning the unit he wants to inspect. The inspector uses a portable device with RFID technology. The system on the device will automatically ask the semantic hub, in this case the LogisticHub, for information about the unit. After the unit is scanned, the semantic hub returns the information about the unit to the portable device. The inspector receives the data about the unit from the semantic hub. The event inspecting starts when the inspector selects to inspect the unit. Here I chose to have the device send a new request asking the semantic hub for inspection data. The semantic hub returns the inspection checklist for the units Z-015 type. If this is the inspection at operator, and not owner, the semantic hub will also return the data from the owners inspection. The inspector receives the inspection data, performs the physical inspection filling out the checklist form, and when the inspection is done sends it to the semantic hub.

Chapter 6

Technologies used to develop a semantic hub

This chapter will explore how the different technologies used in a semantic hub fit with the system development process. Information systems require the use of technologies. These technologies need to be part of the system design and implementation. I want to examine if the technologies used in developing a semantic hub has any effect on the development process. To test these different technologies I have chosen to develop a simple prototype that will utilise certain Semantic Web technologies, and through the process of developing this prototype I will examine how these technologies affect different aspects of the development process. The choice to build a simple prototype of the semantic hub started with the lack of knowledge on access control handling in Semantic Web technologies, and an expressed wish from stakeholders that I examine the security technologies surrounding the semantic hub.

This chapter will first examine the process of serialising and coding the ontology from the techniques used in chapter 5. This to make it possible to implement the ontology into a triplestore. It will then examine the development of the prototype and how the different parts were implemented. It will then examine what technologies could be used to integrate with the semantic hub.

6.1 Serialisation of the Ontology

The ontology was coded using the resource Protégé 4.2 (Protégé, 2013). By taking the UML class diagram and map the classes in that model to the classes in the ontology, and the relations in the UML class diagram to object properties and the class attributes to data properties in the ontology.

Extra events were added that were not modelled into the UML class diagram. Some of the events where initially thought of as states the unit would have during its life cycle. Some of these events, that transpire in the life cycle of the CCU, were added to the class diagram, to give a better understanding of what useful information they could provide. The 'Logistichub Requirements Specification', (PD3) presented a set of business events that would be part of the LogisticHub. I chose to only add those events, those outside that scope are not interesting for this case.

Also, certain attributes, such as `location` that were modelled were seen as possible classes during the writing of the ontology. It became clear through discussion in meetings and interviews that location would be given as a geo-fence, but that these geo-fences could also contain geo-fences making it possible to have location within locations (ME5,IN1). There are also mobile facilities that will have locations that are ever changing and these might be better to tag with a coordinate.

The LogisticHub requirements also introduce another organisations subclass called `custodian`. A custodian is somebody that has the CCU in custody. The custodian has nothing to do in regards to the inspections, but they have rights to events that are triggered while the CCU is in their custody. Therefore they might be custodians while a unit is inspected and would need rights to the inspecting event, but not the inspection data. When following the ontology building methodology there are certain decision that need to be made to make the ontology best fit with what one tries to achieve.

6.1.1 Formal vs Informal

One goal with the semantic hub is to enable interoperability and communication between organisations, but it is also supposed to provide better reports and information for decision making (ME1,ME8). For interoperability and communication between systems you want the ontology to be as formal as possible, but when the ontology is supposed to communicate between people, you want a more informal ontology written in natural language (Uschold, 1996). From an initial meeting with the Statoil stakeholder (ME1,ME2) he expressed that with their systems today it took specialised people to create new queries, and often these queries would cost a lot of money to create, and the process of creating them would take a long time. Therefore the ontology should be created in a semi-formal manner. This to both support easy readability of query results from the triplestore data, as users could more easily write custom queries, and at the same time give the ontology a certain degree of formality to support machine readability. Providing something from both worlds.

6.1.2 Using Existing Ontologies

A recommendation given by Noy and Mcguinness (2000) is to use existing ontologies as much as possible and only modify and extend these to fit into your scope and domain. In oil and gas industry there exists an ISO standard (ISO 15926), that has been made into an OWL[1] ontology. This ontology covers only a small part of the terms and concepts present in this presented case. I also got recommendations in meetings by stakeholders (ME2,ME3) that this ISO 15926 standard is very complex and hard to read and that they did not want me to use to much time focusing on that standard, and that I should rather focus my time on making an ontology to match the given case instead. Later I learned that the 'Logistichub Requirements Specification' states that the LogisticHub ontology will be based on the ISO 15926, but from email correspondence (MA2) I have received statements that mapping this ontology to LogisticHub ontology is not easy task, and very time consuming.

6.2 Creating a prototype

In several of the documents, it was stated that the semantic hub should be accessible from anywhere. This was also stated as a desired feature in meetings with stakeholder (The Norwegian Oil and Gas Association, 2012, UD1,ME2). It was also stated that the semantic hub should be accessible through a Web interface. This made the making

[1] http://www.w3.org/TR/owl-features/

of the prototype interface as a Web application the obvious choice. Making a Web application gives support for the semantic hub to be responsively designed, so it can be used on portable devices, in addition to computers. The Web application was developed in Java Spring utilising the charts from the business process previously described in Section 5.2. The prototype is supported with an OpenRDF Sesame triplestore as data store, back end that used the serialised ontology from Section 6.1. This triplestore was filled with random simulated data created by the researcher in an attempt to create a real experience.

The application was developed over four weeks following an agile methodology of development. Thus developing the prototype in an iterative processes by adding new functionality to the system in every iteration.

I had very little experience with access control handling on semantic data stored in a triplestore. Therefore this became the main focus when exploring the Semantic Web technologies with the prototype. I wanted to test the access control broker, and see how the implementation of the access control broker would effect the development process of the semantic hub. To develop the prototype and the access broker, several different things needed to be implemented and coded in order for it to work:

Triple store: I needed a triplestore to store data. This triple store needed to support named graphs and the TriG notation as well as having a SPARQL 1.1 endpoint to query. I found that Sesame (Aduna, 2012) covered all my needs as well as had a nice Web interface to test write queries.

Access broker The most important part of the prototype. Attempting to recreate the access broker and making it fit the case of the semantic hub. As a development story I had to divide this into several smaller pieces:

Context access graph needed to be created from user information either queried from a triple store or collected from the users device. For this prototype the the context graph was constructed into a query from the login information provided by the user. The result from the query was the users context graph.

Access Policies were created and added to a triple store, as triple statements that state what named graphs it grants access to, and links to what access conditions needs to be fulfilled in order for access to be granted. These access policies will be queried for, and returned to the access broker.

Access conditions sets are collections of access conditions. These access conditions are queries that will in turn be executed, binding the users context access graph. If the ASK query result is true, the named graphs of applied to the access polices gets returned. There also needs to be added handling for disjunctive and conjunctive access conditions.

Safe query is the query that is the result of the access broker. It gets constructed by adding FROM NAMED graphs into the initial query from the user. This safe query gets executed and will give a safe result.

Query: Different queries were created by attempting to recreate different aspects of case problems, as described in Section 5.2. These queries get sent to the access broker and returned safe. When they are returned safe they get executed and the result is returned to the user in a Web interface.

Login session was created to gain information about the user. This information might be stored anywhere. In this prototype I used simple Spring security that used an XML file with stored user names and passwords. This login data were linked to a users graph in the triplestore, that contains statements about that users access roles.

Web interface was created to support simple user interaction. A login screen was created for users to login with their credentials. This was followed by a simple menu screen which contained links and text boxes to predefined queries that could be executed. These queries returned a result of that query to a new Web view if user had access to that information.

6.3 Access control

With the release of the 'Logistichub Requirements Specification', (PD3), organisation stakeholders changed their focus to sharing of data through the semantic hub. They expressed concerns about how the semantic hub and Semantic Web technologies can ensure that their data and privacy issues are handled securely. Both meetings and interviews (ME5,ME7,IN2) revealed that both security and privacy of data is an absolute for many organisations.

I chose an access control framework created by Costabello, Villata et al. (2012), because it promises work as a simple layer on top of any SPARQL endpoint, making it a very flexible solution. It also relies on W3C recommendation, so it does not introduce any new languages to complicate the matter. It also has access control handling for creating, reading, updating, deleting (CRUD) from the triplestore, supporting the use of SPARQL 1.1 UPDATE. Furthermore it supports context aware access control handling, which could be useful for portable devices, as it has been stated by stakeholders that the loss of portable devices might prove to be a security risk (Costabello, Villata et al., 2012, IN3).

The idea behind the the access control framework (2.6.1) is to give roles or users access to query specific graphs. Either to view information, change information or delete information. One must therefore decide ownership of the data. Who owns the information and therefore has the right to share that information as they wish. When the information is shared by graphs, a user will have access to the whole graph if given the rights to it. To give organisations ownership of graphs, we use named graph theory proposed by Carroll et al. (2005a) to assert things about the graph as shown in Figure 6.1(Watkins & Nicole, 2006). Another way of defining ownership of a graph is to make a named graph a subgraph of another named graph as shown in Figure 6.1. This provides the option to define access to graphs all the subgraphs of the supergraph, making the access policies dynamic.

```
@prefix : <http://logistichub.org/> .

:G1 {
:Statoil foaf:name "Statoil"
:Statoil owns:G2
:G3 rdfg:subGraphOf :G1
}
G2: {
:John foaf:name "John"
}
:G3 {
:project1234 rdf:type :Project
}
```

FIGURE 6.1: Named sub graph

One of the desires presented in a meeting was to share inspection documents across different organisations, and still have the possibility to not share the event. To achieve this the simple prototype was developed, in an attempt to serialise the ontology in a triplestore, with triple statements of sample data, and with an access control broker as a security layer on top of the triplestore.

6.3.1 Access Control Broker

The access control framework works as a layer on top of a SPARQL endpoint (Costabello & Villata, 2012). Since any SPARQL endpoint would work I started developing the access control broker. The broker would handle all interaction to and from the triplestore. The first issue that appeared was when the broker needed to differentiate between the different types of queries. Read queries always contain a SELECT, inserting data always starts with an INSERT, and deleting data always starts with a DELETE statement so these were easy to pick out. Unlike SQL SPARQL 1.1 does not introduce an UPDATE statement, but rather manage updating of data by using a DELETE and an INSERT statement. Thus when a query contains both the INSERT and DELETE statements, it will qualify as an UPDATE.

The next issue that appeared was when adding new named graphs to the triplestore, the access broker did not have access policies to handle the new graphs. The access control framework supported that access policies can handle several different named graphs, but this was either by type, or by explicitly writing the named graph into the policy. I made minor modification to the access broker by adding support for a query to represent the named graphs in an access policy. The result of this query is a list of named graphs that users have access rights to. This lowers the demand for access policies, and also enables support for access policies for supergraphs to also apply to its subgraphs as well. This was not previously possible the originally purposed access policies.

All this became possible because the access policies and access conditions are all represented in RDF triple statements, supported by an ontology. Costabello and Villata (2012) claim this is a great advantage in their framework, making the access control both flexible and dynamic. This because you only need to change the access policy in the triplestore in order to change the access rights, and the effect is applied to all access roles. I also think that a semantic hub with many different users and organisations, where the organisations already have their own access control solution implemented locally, could easily connect their access control roles to the semantic hub.

I found one major disadvantage that I did not consider before implementing the access control framework. When adding data, to the triplestore, I had to considered how the data structure is structured, in order to ensure secure sharing of data, with the use of the access control handler.

6.3.2 Data structure issue

When creating the triples statement to be added to the triplestore and testing the queries in an attempt to receive a desired result, I found that I had missed, an important part of the data storing, when using an access control handler. I had not considered how the data needed to be stored in order for the right organisations to be able to share only that the information they wanted to share. The ontology that was created only gave meaning about the triple statements that is stored in the triplestore, and did not consider the need for restricted access to these statements (Chapter 5).

The access control framework Costabello, Villata et al. (2012) grants and denies access to graphs of statements, and these graphs are named with an URI, and it is only the statements within the specific named graph, that the access broker grants the user access to. Thus all the information that should not be accessed by all should be in different graphs. Creating the need to identify what the different organisations want to share, and what they do not want to share. Sharing of information was discussed in several meetings (ME5,ME7). These meetings gave no clear answer to what the different organisations wanted to share, so I attempted to solve one of the more complex issued presented at these meetings. Namely, hat some organisations wanted to share inspection, but only the information from the document not the event itself (ME6,ME7).

The data that the organisations wants or needs to share with other organisations, will need to mirror the named graphs, so that each named graph contains the triple statements of information that is desired to share. The Norwegian Oil and Gas Association (2012, (PD3)) states data access as access to events, thus the events statements become named graphs as well as triple statements. This way it is possible to give and restrict access to all information contained within the event. Similarly, the document like inspection becomes a named graph, containing triple statements about the inspection document(Appendix A), giving the option to only share the result from an inspection, and not all the information about the inspecting event itself. An example of a named graph can be seen in the Figure 6.2. This graph is an inspecting event containing triples about inspecting event and link to the inspection document, which also is a named graph containing the inspection document in form of triple statements. This gives the option to grant access rights on the inspection document, and still restrict access to the inspection event.

```
lho:event_1361881363 {
    lho:event_1361881363 a lho:Inspecting ;
        lho:lat "60.411" ;
        lho:long "5.011" ;
        lho:eventTimeStamp "2013-02-26 01:22:43"^^xsd:dateTime;
        lho:eventTriggeredBy lho:Statoil .
    lho:document_1361881363 a lho:Inspection ;
    }
```

FIGURE 6.2: Example of an named graph of an event in TriG notation

6.4 Serialise data

To provide a realistic base to test the prototype, I needed a triplestore to supply the prototype with data. I discovered during the development of access broker (6.3.1), that I needed a triplestore to support serialisation of named graphs in TriG[2] notation. Because of this I chose to use the triplestore OpenRDF Sesame[3]. This triplestore is an open tool with a SPARQL endpoint and a good visual GUI to test queries and review data.

I populated the triplestore with simulated data, created to be similar to how the real data might be represented. Since the LogisticHub is a system with new information, I had to rely on archive records from Statoil (UD2,UD3) to fill the inspection events and simulate the rest. When simulating data the interesting question became evident. Should event graphs and the event triple be the same URI or should the graph have its own URI? Since the ontology was built without the thought of using named graphs this was not previously considered. I tried both options when testing the prototype, and

[2]http://www.w3.org/2010/01/Turtle/Trig
[3]http://www.openrdf.org

landed on it being less confusing and less ambiguous using the same URI for the event graph and the graph. In Figure 6.2 you see an example of a graph event of the type `lho:Inspecting`. This event has relevant information about the event and a document which also is a graph.

This graph structure creates another complexity, that named graphs like document is here a subgraph of the eventgraph. I investigated, and found that a way to do this was by adding a triple statement in the subgraph, telling what supergraph it belongs to. Thus access rights can be set on the document graph, as well as the event graph. If done the other way around, the user adding the graph would then also need rights to add information to the supergraph as well as the subgraph, and that might not be viable in all situations.

A big challenge for the semantic hub on the Web is distinguishing whether the data that is published is reliable or not. Through a meeting there did not seem to be any concern about the reliability of the data. One Statoil stakeholder said during an interview (ME1) that data reliability should have been considered in the data model, and that since most of the data would be published from known sources he did not see this as a problem. The RDF triples do not support statements about the triples themselves, but with the use of named graphs, one can create statements about a graph. Thus one can create statements about that graph stating who created it, and any other information one would find useful to include. In another meeting (ME3) questions were raised about which timestamp to use, system time or real time, and how this should be handled by the portable device. This seemed to be a technical problem, that could be solved by having the real time as a statement in the in the event graph, and the system time registered as a statement on the named graph (Carroll, Bizer, Hayes & Stickler, 2005b).

Another issue when organisations need to implement Semantic Web technologies, is that they need to consider how much data needs to be converted from old legacy systems to the new semantic hub (Alani, Kalfoglou, O'Hara & Shadbolt, 2005). For the LogisticHub this is for the most part only the master data that, already exists in the existing systems and needs to be migrated over. Since LogisticHub track and trace tasks are not covered by existing systems, all the new event data, being created by the LogisitcHub will be added as it is created (IN1,IN2). When it comes to inspection data at least Statoil has a system for storing this information, but mapping that to RDF and a triplstore should be an easy migration, because this data is a relative small collection (ME2,UD2,UD3).

6.5 Query building & SPARQL 1.1

Since the access control broker promises to work on all SPARQL endpoints, made SPARQL the language in which I chose to query the triplestore. I needed to build queries to match the tasks presented in the BPMN models (5.2), so they would return the data needed by the task, and only the data the user has access to. The more elaborate the task, the more complex the queries became. It also became evident to that the complexity increased when the triples I wanted were stored in different named graphs. In order to get all the statements needed, I had to know which statements were in stored in which graph. Figure 6.3 shows an example query getting all the inspections that one user has access to. All the FROM NAMED ensures that only results that `lho:john` has access to are returned. This query shows that it can be more demanding for users creating custom queries themselves, since they not only need to know the query language, but also the data structure the statements are stores in, to get the query results they want. I attempted to use FROM instead of FROM NAMED, but that only gave incorrect results.

Further findings show that the modified query sent by the access broker, will quickly get huge amounts of FROM NAMED graphs in the safe query. This could effect the response time, quickly becoming unbearable for a system like this. Making the semantic hub seem as lower quality and give lower user satisfaction. Then again, one could always add more machine power to solve the problem, which is cheap in today's world (Costabello, Villata et al., 2012).

6.5.1 SPARQL REST

One thought behind the LogisicHub is for organisations to connect to the hub via a Web service. The Web service in LogisticHub is set to be either a SOAP[4] and or a REST[5] where XML schemas will be used to pass information to and from the Web service. The

[4]http://searchsoa.techtarget.com/definition/SOAP
[5]http://searchsoa.techtarget.com/definition/REST

```
SELECT DISTINCT  ?event ?time ?inspection ?unit ?org ?orgName ?project
FROM NAMED lho:Halliburton
FROM NAMED lho:event_1361963009
FROM NAMED lho:unit76999
FROM NAMED lho:unit76556
FROM NAMED lho:CCB
FROM NAMED lho:z015Graph
FROM NAMED lho:Shell
FROM NAMED lho:document_1361963009
FROM NAMED lho:hiring15464
FROM NAMED lho:hiring1
FROM NAMED lho:project_1234
FROM NAMED lho:Statoil
FROM NAMED lho:event_1361881363
FROM NAMED lho:event_1361958858
FROM NAMED lho:unit98765
FROM NAMED lho:unit12554
FROM NAMED lho:unit12345
FROM NAMED lho:Schlumberger
FROM NAMED lho:unit23456
FROM NAMED lho:document_1361958858
FROM NAMED lho:document_1361881363
WHERE
  { GRAPH ?g
      { ?event rdf:type lho:Inspecting .
        ?event lho:eventTimeStamp ?time .
        ?event rdfg:subGraphOf ?project
      }
    GRAPH ?event
      { ?inspection rdf:type lho:Inspection .
        ?inspection lho:inspectedUnit ?unit .
        ?inspection lho:inspectedFor ?org
      }
    GRAPH ?h
      { ?org foaf:name ?orgName }
  }
```

FIGURE 6.3: Example of an SPARQL query that returns a list of all inspections that user `lho:john` has access to view, along with what unit, when and for who it was done

people behind the LogisticHub did not yet know this XML would be formatted, and when asked if it would be RDF/XML they said it would be up to the developers of the LogisticHub to chose that (IN3). Other stakeholders stated that they would like to use JSON[6] for sending data and communicating with the semantic hub (ME5,ME9). Since SPARQL 1.1 is not only a query language, but also a protocol, a SPARQL endpoint can easily be adapted into a REST Web service, providing a service that organisations can

[6]http://www.json.org

utilise to get the information they want, and since many of these endpoints support both JSON and XML, it will be up to the organisations themselves, which format they chose to communicate with the semantic hub, as long as it is in RDF format. Also, since the access control handler is built to fit to any SPARQL endpoint and it supports SPARQL 1.1 UPADATE, the Web service can also provided with access control handling.

Chapter 7

Using the LogisticHub

This chapter will first examine the benefits of the semantic hub as seen by organisations. These benefits are very important for the initial phase of system development. This is because these benefits can be decisive for whether or not a new system will be developed. With a semantic hub, there are many different organisations and stakeholders involved in the process. Examining the benefits should create a better understanding of why they are developing a semantic hub.

The second part in this chapter will explore the intended use of the semantic hub by the organisations. Due to the special nature of the semantic hub, in being a centre for business activities in the offshore supply chain, and connecting several different organisations and their systems together. How the organisation intends to use the semantic hub, might effect how it will be developed.

To investigate these two parts I have collected data from interviews, meetings and some different types of documentation (3.4). The LogisticHub has only recently been approved for full development, therefore the plans for how the different organisation see the semantic hub have changed over time.

The benefits and use are part of the initial stages of the system development process. However, this chapter will be at the end of this paper. This is because this case study is written in a chronological manner, and a more detailed inquiry into the benefits and intended use was done after the examination of development techniques and technologies.

7.1 Desired Benefits to Desired Outcome

The LogisicHub is being created to provide organisations with a track and trace solution for the CCUs they either own, rent or have in their custody. One of the main long term goals of the track and trace solution, is to give the organisations more and better information to be used in decision making. The organisations stakeholders do not only want better information, but also want the semantic hub to give them this information in real time, in order to use this information in planning and decision making in the day to day operation. A given example of how this might work is, a CCU that is supposed to be shipped are sent to the oil rig TrollA, has just arrived at gate. The correct personnel get sent information that the CCU has arrived, so this CCU can be included for the day's shipping to TrollA. They also want to use the larger amount of better quality data to analyse and use in future planning and decision making at a higher level. Examples of what the organisations think might beneficial are movement patterns of the CCUs, how long a CCU had been standing still in one place, and also how many CCUs are now at this location (ME1,ME2,ME8,IN2,PD3).

To provide the best possible information, the semantic hub will collect information from several different sources and organisations, to become a hub of sharing. Then the data is exchanged between the different organisations. This kind of information sharing hardly exists in the offshore supply chain today (IN2). The sharing of data between the organisations provides them with more information they did not previously have access to, and should provide new possibilities, for example gaining more useful data at an earlier point in time. Another example of this might be to receive information about when the CCU departed from the owners location, with an estimated arrival time. This data can then be sent or collected automatically by the receiving organisation, and adding it to their local system for better planning. This will creating a much faster information flow than what exists today. This should enable better decision making in the daily operations, since the people responsible can get more information from outside the organisation. This streamlining alone is estimated by the stakeholder to cause huge cost reductions in the offshore supply chain (ME3.ME8,IN3).

Initially, and before the release of 'Logistichub Requirements Specification', (PD3) Statoil stakeholder, stated that they could see benefits by sharing inspection data for CCUs of type Z-015 (ME1). They expressed that by sharing inspections with Z-015 owners, they could streamline the complex process of inspection required by law. The streamlining would be enabled by making the inspection digital, where all the information that needs to be sent to and from hirer and owner would be sent through the semantic hub, and not by paper, like today. Also, by providing more automated tasks and messaging, as described in Sections 5.2.2 and 5.2.1, the streamlining would be furthered. Furthermore, the Statoil stakeholder also wanted to share the inspection data with other competing organisation to provide a larger set of data to be used for equipment fault checking. From interviews, and meetings (ME2,ME8,IN2) it was expressed that the subject of sharing inspection was not openly discussed yet, and that this needed to be done before anything could be agreed upon. I would recommend starting an official discussion on the subject, since feedback on the subject has become more positive during the duration of the research, as more stakeholders has come to see the advantages that were presented (IN2,ME8). With the use of ontologies and several integrated data sources the organisation might draw unseen advantages from inferences that could previously not been seen (Alani et al., 2005, ME8,ME2).

7.2 Communication and data gathering

With a desire to share relevant knowledge between the different organisations in order to reach a desired outcome, Statoil stated in a document (UD1) that they wanted the system to contain as little information as possible and that the information that was already accessible elsewhere should only be linked to the semantic hub. Since most of the legacy data is not available in RDF format, the plan is to collect the master data automatically from available sources (UD5), for example NPD fact pages, PCA RDS, GS1 GLN and the Brønnøysund register, as shown in Figure 7.1. This will be the responsibility of EPIM (PD3). Other master data that is not openly accessible, like CCU master data will be added to the LogsticHub by the owners of the CCUs. Similarly, other organisations will be responsible for adding their own information, which is not openly available for the LogisticHub to automatically update.

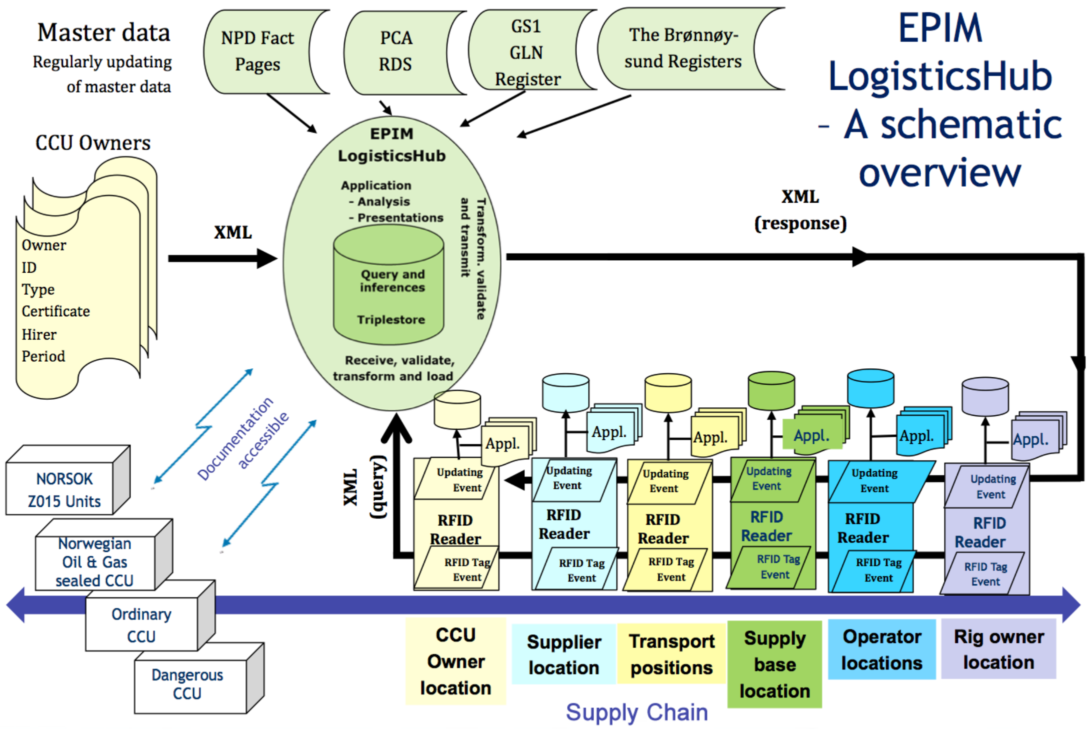

FIGURE 7.1: EPIM's schematic overview of LogisticHub (The Norwegian Oil and Gas Association, 2012)

The owners of CCUs are responsible for providing the correct documentation for the CCUs before it is sent to hire. Unfortunately many of these documents are today only available as human readable digital documents, for example PDF. I have in Chapter 5 examined how one can add parts of these documents into the semantic hub as part of an ontology. I think that there are documents that can be equally beneficial if represented in the ontology. An example might be the hook-up information documents for every facility offshore (Norsk Standard, 2013, PD1). These documents are today displayed in several different manners. In meetings, a Statoil stakeholder expressed that these documents were a very important part of the Z-015 routines, similar to the inspection documents (ME2). These hook-up documents are not described in the 'Logistichub Requirements Specification', but has been expressed as desirable to have included in the LogisticHub. I do not know how much the inspectors use the Z-015 specification document (PD1) during the inspection itself, but the document contains a lot of specific information about certain types of Z-015, and when the inspection is performed digitally on a portable device, making it easy to look up on the device can only be beneficial.

It was also stated in a meeting (ME9) that some organisations might desire to have their inspection data stored locally and that the LogisticHub only linked to it. I think this can be achieved, but only if they store it locally following the ontology set for the semantic hub. If not, one will end up with different organisation storing the inspection data in different formats from each other, and thus disabling the benefit of sharing inspection data as explained in previous sections (5.2.2).

There has also been expressed benefits in removing excessive paper work and making these processes more digital. When shipping equipment like the Z-015, it is required to ship along a lot of different papers and documents. The organisational stakeholders would like to have these documents added to the CCU digitally instead of in an envelope of paper. These documents could be linked directly to the CCU or to the CCU's hired event, enabling that the correct documents are always accessible with the CCU (ME2).

EPIM has expresses a desire to create several different semantic hubs, one at the time, and that they desire to use these as one large knowledge base (UD4). EPIM has today already developed two other semantic hubs and implemented these into the Norwegian oil industry, and stakeholders from EPIM stated in interview that they are already planning another semantic hub (IN3).

During the life cycle of a CCU, the different organisations that is handling the CCU will trigger events about the CCU, and send the event data to the LogisticHub. I have been told in a meeting (ME6) that when an organisation sends the event data to the LogisticHub, the same data that will be sent to them in return. I presume that this is the same data in that the semantic hub gets, and that the organisation only gets sent the data they themselves triggered, or possibly the data that is trigger for them by a sub-contractor organisation. One stakeholder stated in an interview (IN2) that he wanted to use this data to build their own data store of events. This was because he wanted to reduce communication being sent to and from the LogisticHub. I think this would partially work against the EPIM goal of creating several knowledge bases to be queried at the same time. An interview with an EPIM stakeholder (IN3) revealed that they believe that organisations only need to experience the use of the semantic hub and the Semantic technology in order to appreciate them. The organisations that are still reluctant to share knowledge through the semantic hub will change their opinion on sharing as they discover that the benefits are greater than the risk of sharing (IN3) From interviews and meetings there has been stated that organisation would like to integrate the LogisticHub as much as possible to their own systems.

7.3 Shadow System

The semantic hub will provide a web service the organisations can use to integrate the LogisticHub to their own systems. The interviews revealed that one organisation is thinking of building a shadow system. The shadow system will be an event management system that tracks all events in their supply chain. This module is supposes to be an extension to their existing ERP system. The event manager is supposes to capture the same information that is going to the LogisticHub, but the shadow system will also be integrated to the organisations internal data (IN1,ME7,ME6). When the organisations stakeholder was asked if they were considering using Semantic Web technologies to build their local shadow system, the answer was *"we are looking at it, but most likely not"* (IN1).

One stakeholder said he wanted all the information they generate in the LogisticHub stored in a local data storage. This because he thought that they would otherwise generate to much traffic to and from the LogisticHub. He was not sure what technology they would use, but sure that it would be part of their SAP system (IN2).

These statements make me presume that some of the usefulness of the semantic hub will disappear. It was not clear how they would handle and store the information they only have access to in the LogisticHub, but have not generated themselves. How will they implement this information into their shadow system or local data store? They could query for it, and store the query result in their local system regularly. This can lower the generated traffic, but will also lower the information quality. In addition, by not using Semantic Web technologies to store their information locally, they will lose a lot of meaning in the data if they want to share it in the future. One early expressed desire by a stakeholder was to use Semantic Web technologies to help the organisation with internal interoperability as well as external interoperability (ME1). This desire seems to have been misplaced in the process. Alani et al. (2005) states that the benefits for organisations in putting their data into RDF, becomes more apparent as publish in data in RDF format.

When not utilising the Semantic Web technologies in the shadow system and still following the old legacy ways similar to what is already being used in their existing systems, which as of today is not capable of providing an interoperable solution. Organisation will not be provided with the best highest gains from the benefits of the semantic hub.

There also seem to be some disagreement on how much of the semantic hub the organisation will utilise. Some organisations seem reluctant to include certain parts of the systems business events. The organisations state that only the data that is generated to and from their organisation will be stored in the LogisticHub on their part. Still it seems that the business processes they do not want part of the LogisticHub, are still interesting to have in a local system. This make me think that it is not the data itself that is not interesting, but the thought of storing it on the LogisticHub (IN1,IN2). This shows that having a good access control framework is very important for a semantic hub in an enterprise environment. This reluctance could also be because of the way stakeholders view the LogisticHub. Interviews seem to indicate that some organisations think that all the data stored in the LogisticHub, is automatically shared with someone. Maybe they do not see the organisation EPIM as a part of their organisation even though they are part owner or there may be a natural lack of trust within the industry (IN1,IN2).

I can understand that organisations would like total control of their own information. Especially when they do not need to share that information, and thus do not really need to store in a joint data store. If the organisation should be able to pick and chose from the functionality supported by the LogisticHub this could affect the commitment given to the semantic hub by the organisations. This might not have a big effect on LogisticHub, but could put strain on how data is gathered for the semantic hub. Giving the semantic hub many special cases could complicate data gathering.

Chapter 8

Discussion

This thesis has changed in perspective throughout the research, but there has always been one common underlying goal of examining a semantic hub being developed in an enterprise environment. The question I ended up attempting to answer was: How is the development process for a semantic hub compared to the development of other information system? To examine this question more deeply, I created four sub-questions that could help answer the main question.

In this chapter, I will first discuss the findings in view of the sub-questions, and then sum up these sub-questions to answer the main question on this thesis. In the second part of this chapter I will present some other lessons learned from the research that are relevant for semantic hub development, but do not fit entirely with the research question. In the end of this chapter I will discuss the research process and the research method used to answer these research questions.

8.1 The Questions

8.1.1 Research Question 1

New desired benefits are often what starts the creation of new information systems. Benefits need to be found to trigger deeper investigations and analyses for developing a new information system. Benefits can often be measured in clear value as cost reduction, but this is hard to calculate prior to the implementation of a new system. Thus often the benefits that are set for planning a system are either perceived or desired benefits. The benefits are also important when planning to develop a system like a semantic hub. We therefore want to find: **What benefits are seen by the organisation as desired outcomes from the semantic hub?**

It is difficult to measure benefits before the information system has been implemented, but it needs to be part of the early stages of the development process of the information system. Thus all benefits perceived before they are measured can be seen as desired benefits. These benefits are what shapes and drives the way to the final outcome as perceived by the different organisations stakeholders. The main desired benefit is an overall cost reduction in the offshore supply chain. This will be achieved by providing a track and trace system for CCUs in the offshore supply chain. The semantic hub LogisticHub will use RFID technology to capture data from several different organisations, and this information will be shared across the different organisational systems. The information gathered by the system is supposed to give the organisations insight into the patterns of the supply chain, and thus improve efficiency and productivity. This is estimated to reduce the overall costs in the supply chain by a large amount (The Norwegian Oil and Gas Association, 2012, lh).

The semantic hub will give the organisations more data about the different events happening in the supply chain. This to give the users more data, that can be analysed and used in decision making. An example of what might be useful, is to know how many CCUs have been in holding for how long (IN1). Information is shared between the organisations to gain data about CCU movements outside the organisations. The organisations want this information to be available for its users in real time. This to improve the day to day operations in both planning and decision making, as for example when a CCU arrives at gate, the right people will get notified about this event, so they can decide if this CCU should be added to today's shipping (ME8,ME6).

The semantic hub is also supposed to give functionality that will make it possible for organisations to streamline business processes through automation in handling of tasks. This can be achieved by sharing data between organisations. Inspections shared between hirer and owner of CCUs enable the possibility for faster and more efficient handling of the inspecting tasks. It has been stated that streamlining the tasks of handling Z-015 temporary equipment, could provide a large cost reduction. The Z-015 temporary equipment is very expensive to rent, sometimes 100000 NOK a day. Even short delays might bring up a lot of unnecessary costs (ME2). Statoil stakeholder said in a interview that 100 000 here and 100 000 there quickly becomes a lot of money (IN1).

Another desired benefit with the semantic hub was to reducing paperwork and menial tasks, as for example turning inspection forms in paper into digital records. Today all inspections are done on paper and added digitally to a data store, but by streamlining the process into the semantic hub, it would become all digital. The inspectors can collect the forms needed for the inspection digitally from the semantic hub, and then when they are done with the inspection, they can send it back digitally. This is explained in Section 5.2.2 (ME2,PD1).

Most of this data needs to be stored in a joint data store, thus the organisations want the semantic hub to provide a secure solution for storing the data (ME6,IN3). The organisation needs to be sure that their data privacy concerns are handled correctly, so only those users and organisations that should have access to the data, have access to it. One can understand their wish to protect their own interests, but this will might effect the use of the semantic hub or the way the data is handled or stored. Not feeling secure will drive the organisations further towards their internal systems, and only communicate the data which is absolutely needed in order to fulfil their responsibilities (ME6,ME8).

The possibility to include inspections and sharing these needed documents has triggered a new ideas from the stakeholders. They have expressed interest in including the tasks like yearly certification renewals for various equipment and CCUs, in a similar way as purposed with the Z-015, and thus removing even more paperwork and reducing cost even further, extending the possible use of the semantic hub (ME9).

The joint organisations that are responsible for the LogisticHub aims to create a large knowledge base by building several different semantic hubs, one at the time. These hubs are going to cover different aspects of the oil and gas industry, and with time should bring the industry a very large source of knowledge that will become very valuable for its users (IN3,PD3). A Statoil stakeholder expressed that he desired possibilities creating custom queries without having to hire expert developers to implement them, having to wait for several months and spend lots of money, before getting the results (ME2).

Organisations sometimes seem to only look for benefits from sharing data with other organisations that they have a close working relationship with, and do not seem to consider the benefits from sharing data beyond these close connections. This might bring less creativity to how the semantic hub could best be utilised, and what benefits that can be envisioned from the system. This can be seen in the interviews, where one stakeholder wants to share inspections of Z-015 with all organisations, to gain better fault tracking of this equipment, but a stakeholder in another organisation did not see the same value at first, but later as more was explained became more interested (IN1,IN2,ME2,ME4). This makes me think that the benefits can be seen as the extent in which the information system contributes to the success of the individuals, groups, organisations and nations (Petter, DeLone & McLean, 2008).

8.1.2 Research Question 2

There are many desired outcomes perceived from the LogisticHub. Many different organisations have different opinions of how the semantic hub is best to be utilised, for the organisations to reach their long term goals. To do this the semantic hub will need to facilitate all the different organisations desires. The semantic hub will provide a solution that is not possible with existing systems. Not only do infrastructure and hardware like RFID scanners and RFID chips need to be installed in equipment, and at different installations. The different systems used by the different organisations are insufficient to mediate this new information. To better understand how the semantic hub fits into this, I have chosen to examine: **How do the organisations intended to use the semantic hub?** This in order to examine how intended use might guide the development of the semantic hub.

The main use for the LogisticHub is to encapsulate business processes that the organisations think are handled badly today in the offshore supply chain. The semantic hub will capture events triggered by CCUs in the offshore supply chain and log these events in a semantic data store. Initially, the view of the research was on how the semantic hub would handle the business processes related to the CCUs of the type Z-015, and with a special focus on the tasks surrounding inspecting. The idea was to make the task handling of Z-015 better both for external and internal systems. This because the system that handle this today is not integrated with the organisations internal ERP system, and was originally built as a temporary solution. Making the semantic hub a solution for adapting these tasks, either as a service where the tasks are handled in the local system, or handle were the tasks are handled through a web interface from the semantic hub. By handling the tasks directly on the semantic hub you can provide the inspectors with a common shared tool to use when inspecting Z-015 units (ME2,UD1,UD2).

Some of the larger operators will build their own systems that are independent from the semantic hub, but will connect to the semantic hubs web service to get the data it needs. They also wanted to include more internal private data that is not stored in the semantic hub, providing them with a detailed tracking information that also includes internal movement. This system will be built as a module for their existing system. A stakeholder from the organisation stated that this module would unlikely be built using Semantic Web technologies, but that they were considering it. By creating a local system using Semantic Web technologies, the system could then link the data from LogisticHub to their local data. This will better ensure that the organisation s the information quality they want, and does not lose any information from unnecessary data translations, from a richer to a lesser data model. Additionally this can be said to be the first step towards the organisation having created their own knowledge base, and with this knowledge base they can add all the internal data, that they do not desire to share, giving them an even a greater source of data (IN1,IN2). This could push the development of the ontology to be extended to also include internal business processes as well as external. Giving an extra push towards using and utilising the semantic hub.

One organisation wanted to store as much data as possible locally. Using the same data they send to the LogisticHub and storing it in a local data store. This was because they did not want to generate to much traffic to and from the LogisticHub. They will still need to use the semantic hubs web service regularly to ensure that the data stored locally is complete, to not lose any quality of information. This because some of the information shared on the semantic hub will be owned, and triggered, by other organisations, and they will need to collect this data from the semantic hub web service. One way to fully utilise the local data store would be to create it as a triplestore, but that was expressed as being unlikely by the stakeholder (IN1).

There has also been triggered other ideas for how the semantic hub can be utilised during this research. CCU owners expressed that they wanted to add yearly certification renewal for the CCUs as part of the semantic hub. This to both reduce the paperwork in this process, and to improve the sharing of the certificate data that is mandatory to follow the CCU. The handling of this would be similar to the handling of inspections of Z-015, explained in Chapter 5 (ME9,PD3).

Many organisations have different ideas on how to integrate the hub. In the development process, these ideas needs to be straightened out. Either all organisations must agree on how it should be used, which seems hard since one problem in this project has been to agree on common standards, or that the semantic hub needs to facilitate several different scenarios, where for example data is only stored for the cases that the organisations wish it to be stored. This will put extra strain on the development of the system, but might help ensure that it reaches the long term goals for all organisations.

The importance of the intended use is to identify how the information system is perceived to reach its desired goals. Petter et al. (2008) states that there is a close relationship between the intended use and actual use of an information system and that support for the use of the system leads to benefits. Thus the development process needs to include investigation and analysis into the intended use of the semantic hub in order to make sure it can fulfil these.

8.1.3 Research Question 3

Which development techniques could fit with the development process of a semantic hub? The development techniques fitted well in the development process. Statoil stakeholder states that the ontology is the foundation of the LogisticHub (IN1). When building semantic systems you might be required to follow certain techniques, as for example ontology building methodologies for creating an ontology. These methods uses techniques like competency questions created from scenarios and system context stories, to best ensure that the right domain scope has been selected for the semantic hub (Noy & Mcguinness, 2000).

The techniques worked to combine the business part of the development process together with the more technical development part. The UML class diagram fitted well with the ontology building techniques. Since the UML class diagrams concepts can easily be adopted to fit with the ontology building concepts. In addition, UML is a well known and established open source technique used in system development methodologies, and many developers have good knowledge on how to use it. It creates a platform for discussing the systems domain with its stakeholders, that often are more knowledgeable in the domain. This provides a basis for organising the different parts of the domain, like the ontology's terms, concepts, attributes and relations in a visual model that can more easily be communicated (Fowler, 2004).

A BPMN model was used to investigate the business tasks and in mapping the desired outcomes. This made it easier to identify what needed to be created in order to achieve the systems desired goals, examples of this might be, what messages get sent automatically, and what data needs to be shared across systems (5.2.1). Even though some of these goals can be seen as internal improvements for efficiency and productivity, these benefits could not have been achieved without the cross communication handled by the semantic hub. This shows the value of the semantic hub how it becomes a link between the systems.

BPMN worked also as a cross checker to ensure that the correct domain and scope were captured. It provided support for matching the competency questions against the BPMN model. One way was to use the data flow in the BPMN model to validate the ontology like a competency question. The other way was to use the competency question to validate the data flow in the model. This provided a way of evaluating the ontology, making sure it got the information that the system requires, and that the system got all the functionality mapped in the model (Uschold, 1996; Noy & Mcguinness, 2000).

Gable, Sedera and Chan (2008) claim that by following well grounded methodologies in the development of systems, one increases the chances for the system to reach it's goals. In one of the interviews, the subject stated that there seems to be a general lack of knowledge in creating and developing a Semantic Web technology application like the LogisticHub (IN3). Thus using techniques from well grounded development methods, that should be known by developers, will provide a great advantage. Specially when some of the techniques can be used to cross check with techniques that are lesser known and required in the development process.

8.1.4 Research Question 4

How do the technologies used in a semantic hub have impact on the development process?

I have chosen to examine this by looking at the example of access control handling on Semantic Web technologies. The access control framework I chose to implement in the prototype requirements depends on the use of certain technologies. Already, the semantic hub requires the use certain Semantic Web technologies, and these choices are further limited by the access control framework. These limitations make it more difficult for the developers when choosing what technologies to use, and can change how the semantic hub will be designed and implemented.

The development techniques are used to ensure the thoroughness of the development process, but you need to chose the right techniques in order to ensure that all the business processes are covered. During the development process of implementing the access control framework into the prototype, I discovered deficiency in how the data was structured in the triplestore. When using Semantic Web technologies you store the data in form of triple statements in a graph store. The access control framework works by giving users access rights to named graphs. Thus, the triple statements needing different types of restrictions, needed to be stored in different named graphs. This was not considered, and caused a setback in the development process. If forced me to go back and review the sharing and ownership of data (4.4). This was a direct result from choosing an access control framework. By choosing a different access control framework, this delay in the development might not have been necessary, but another access control framework might have had other considerations that needed to be taken into account.

The access control framework uses triple statements combined with an ontology for setting and assigning the access control roles and rights. This solution enables the option for organisations to connect their local access control roles to the access control statements, and thus when assigning access control to users locally this can also be updated dynamically on the semantic hub. This feature might require the development of a new or extended ontology, to match the security handing from the industry to the semantic hub. This creates a need to review the access conditions and access rights in the supply chain more thoroughly. Then this information can either be used to build a new ontology or to extend an existing ontology (Costabello, Villata et al., 2012).

When exploring which users should have rights to change the different data, there were no clear definition on this subject. At one point it seemed like there might not even be a need for users to be able to change event data in the semantic hub, and that if changes need to be done this could be handled by administrative personnel at EPIM. This would open up the possibility for a much simpler access control handler, but then what about the future. This is only a starting point for this semantic hub, and since access control framework seems to be dependent on the technologies used, it would be important to find a solution that also support future possibilities (IN3).

Through examining the prototype, I found that the Semantic Web technologies used could easily be ported. The SPARQL 1.1 technology could handle all data sent to and from the semantic hub through its endpoint, and because the access control framework works as a layer on top of any SPARQL endpoint, this data would be secure. This makes the process of finding and selecting technologies that are flexible, such as the access control handler, something that can provide benefits to the development process. Not only because it is so flexible, but also because it would make the external usability much easier for its users. This because they can easily create new queries to get the data they want for their systems, instead of develop a new service for every type of information retrieval scenario the organisation wants possible. This is supported by SPARQL because it is not only a query language for RDF, but also a protocol for retrieving RDF data from a graph store. With this support, the endpoint can easily be integrated as a RESTful web service that also has access control. This makes the development process easier, since one service can handle all the requests (Feigenbaum, Williams, Clark & Torres, 2013).

A desire was expressed to send data formatted in JSON in stead of XML (ME7,ME9). With the SPARQL protocol this is possible. The HTTP request only needs to contain information about what format the user wants to get the result back in. This also makes it easier for the development process, pushing the decision to use XML or JSON to the system developers of the organisation that are going to use the service. The Semantic Web technologies are still relatively new and untested, especially with access control frameworks. This makes them not tested on such a large scale as the LogisticHub, and can be found insufficient to be used. During testing of the access control framework, I found that even at the small scale of the prototype, the querier quickly contained a huge list of FROM NAMED graphs, as shown in the example in Figure 6.3. This, together with the amount of estimated events in the offshore supply chain, and with all these events as their own graphs, this could reduce the performance to an unbearable level. One could possibly solve this by throwing machine power on it, but that would be a lesser solution. This shows that technologies need to be investigated and tested, before they can be approved for us in implementation of the semantic hub. Part of the evaluation of technology feasibility should be performance, as it leads to providing information fast enough for it to be useful.

All this new technology might also create the need for a longer learning curve, adding to the time the development process takes. It has been expressed by a stakeholder that even information system developers lack knowledge on the use of Semantic Web technologies (IN3). Thus the developers need more time to figure out and investigate the technology before they can be used to create the system. This makes both the processes of integrating and using the system take longer time, but the more they use the technologies, the clearer the benefits will be (Alani et al., 2005).

8.1.5 Main Research Question

Through these four sub questions, I have attempted to answer the main research question: **How is the development process of a semantic hub compared to the development process of other information systems?**

The desired benefits and outcomes showed that there are many different stakeholders involved in the process developing a semantic hub, and even though the different stakeholders have basically the same overall desired benefits, they still have very different views on the desired outcomes of the semantic hub.

The benefits seen by the different organisations need to be achievable through the use of the semantic hub. Thus the outcome of the semantic hub needs to reflect what the different stakeholders want, and what benefits they perceive from using it. In the development of regular information system you investigate deficiencies in the existing systems to see if these can be fixed with the new system. Because the LogisticHub provides a new solution that do not exist in the old systems, making it important to find out how the semantic hub system is perceived more important. Avison and Fitzgerald (2006) recommend that all stakeholders are part of the development process, but in a project like the LogisticHub this is very large group, and it would quickly become expensive and time consuming to include them all. This makes it important to include at leased one representative of each different type of stakeholders are represented in the development process. This to ensure that the system captures all the different aspects desired.

The intended use by the different stakeholders show that many of the organisations intend to build a subsystem to their existing systems that will utilise the semantic hub. Some organisations also consider using a local storage to store the same data as they have access to in the semantic hub. This because they want to lower the traffic to and from the semantic hub. The LogisticHub proposes several different business events that are supposed to be handled by the semantic hub, but organisations have stated that they do not want to use all these events, and that they are only interested in events that happen to and from their organisation. This even though these internal events are interesting for use in the internal subsystem. This can create strain on the development process as the semantic hub might need to support special cases in handling business processes for organisations that do not agree with the systems whole.

Development techniques like UML and BPMN, work well in the process of developing a semantic hub. Both UML and BPMN worked well as supplements to ontology building techniques like competency questions. Additionally they brought extra depth to the development investigation and analysis. When developing for organisations, especially when the stakeholders are involved, and they are not very technical in the area of system development, it is important to be able to use techniques like these to lower this technical barrier, so one can best possible communicate the business domain (Avison & Fitzgerald, 2006).

Certain Semantic Web technologies require the use of certain development techniques. For example, if you need to develop an ontology, to use with your semantic hub, you also need to use techniques that are used in the ontology building methodologies. As with sharing of data where the access control framework might require the use of a development technique for making sure the data structure is correct. Techniques used in developing other information systems can be easily adopted and fitted to the development of the semantic hub.

Semantic Web technologies are one of the few technologies that can make the desired outcomes possible for all the different organisations, because EPIM aims to create one large knowledge base for the whole industry domain. These technologies are for the most part untested in an enterprise environment, making the process of development extra difficult, as the developers might not be sure of how the technologies work, and how they best can utilise them. This creates extra strain on the developers, and they might need to test the technologies before using them to implement the system. This is to make sure that the technology is viable to use, and to see if there are parts of the system that needs to be examined more closely. Choosing the wrong technology to develop parts of the semantic hub can affect the system, as for example with the access control framework, that might not be scalable. These types of issues can result in delays, or in worst case, cause the whole project to fail. This shows that a lack of planning can lead to unforeseen costs with the development of the system and can ultimately lead to abandonment of the system (Avison & Fitzgerald, 2006).

8.2 Other lesson learned

During the research of this case there has been many interesting experiences. Some of these experiences are not directly connected to the development process of the semantic hub, and have more impact on the system itself, but these lessons can be important to have in mind when developing a semantic hub. This because some of these experiences could also occur in the development of other semantic hubs.

8.2.1 Other lessons from technology

Named graphs bring the possibility to create statements about the graphs them selves, an example might be to note on a graph when or who changed it. This provides a solution for adding provenance and trust to the graphs in the semantic hub. When interview subjects were asked how provenance was handled in the system, they stated that it was not being handled by the data model and that they did not considered provenance as a large issue (IN1.IN3).

One organisation stated in a meeting that they wanted to be able to quickly create new queries without it being too expansive. By giving them access to a SPARQL endpoint, users might easily create new queries themselves. Initially, the LogisticHub will only provide predefined queries or queries that are created by selecting from a predefined set of parameters, but a future aim is to give the users the ability to create custom queries. To enable custom queries will provide more possibilities, since the users themselves can then tailor the information they desire, instead of an overall best fit. Custom queries might introduce a higher learning curve from system users, but with both solutions in place, the time the learning would take for customisable queries would be less demanding. They do not have to sacrifice information quality by not providing users with custom queries right away. I would not recommend sacrificing the option for custom queries, which should bring a higher gain to both use and usability of the semantic hub, for a longer learning curve. One could argue that users will not miss what they do not have, but a major part of the Semantic Web technologies is it query power. This power needs to be showed to the organisation for them to experience this, and hopefully implement this technology into their own systems, and thus making a greater impact on the overall benefits from this project (IN3).

This supported with an ontology written in natural language will enable the users to more easily understand the query results. Thus making it possible for them to read the results without having the need for it to be modified for readability. For users to query themselves they would need to learn how to use the query language. This will require the semantic hub to provide users with support for this. The support could be provided by the EPIM, but if querying should be available for all users this could become a large burden.

8.2.2 Lessons Learned from Benefits and Use

In order for the semantic hub to provide shared information that the different organisations can communicate, the different organisations must agree upon a common shared standard of terms and concepts. In an interview (IN1) with a stakeholder responsible for working out theses agreements, he stated that to come together and agree upon a common standard was one of the biggest challenges with the project, but he was positive and said they where 60-70% in agreement (IN2,ME2,PD3).

A Stakeholder stated that they share information with thousands of companies world wide using XML (IN2), and was confident that this solution was adequate, but this type of sharing only provides syntactical interoperability. This can be enough when sharing one to one with an organisation, and if the need is great, but to get this done they will need to create new modules that convert the information from each of the organisations sources. Here is where advantage of an ontology is great, it can be used to connect many heterogeneous sources together that otherwise would not have be possible to merge (Uschold, 1996).

To achieve fault tracking with the semantic hub, several things need to be in place. There need to be a ontology for Z-015, and to provide meaning to the inspection data. These detailed inspection documents represented as triple statements will enable the users to query for very detailed information about the inspections. As a hirer you can get very detailed data about fault history on the equipment you are renting and on the owners of it. Also the hirer can share inspection data with other hirer organisations to gain a larger data source for checking fault history of the equipment. There have been raised concerns that the owners of the Z-015 will not agree to this, and thus not be willing to use the semantic hub for this purpose (ME2,ME8,IN1).

8.3 Research Method

To research the questions above, I performed an exploratory case study examining the development process of the LogisticHub. There were several factors that made me chose to use case study as my research method. I had little to no control over the events taking place in the case project I was examining. The project evolved, as new documents were released stating changes or finalising parts of the development process. This made the stakeholders change their focus for desired area of examination, forcing me to also change my focus so I could get the feedback I needed. Yin (2009) states that when the researcher has to change the research design, it is no longer appropriate to continue with the initial research question, and that one should simply start over. As with my research I had to change the research design to fit with my new situation, forcing me to remake research questions and extend data collecting and review data collected. This provided me with several different data sources suitable for a case study (Yin, 2009).

The case study was approached as a single cases, and in the beginning I was only looking at one site. However, as the research changed in perspective, I needed to include other sites. It is recommended in case studies to study multiple cases and not using only a single case. This because multiple cases yield more information, and with several cases you get higher external validity making it easier to generalise the findings. I chose single case study because the case presented a unique opportunity, to follow an actual ongoing development process of a semantic hub in an enterprise environment, which is mentioned by Yin (2009) as an instance when single case studies can be useful. He also states that *the use of a single case studies can show how investigations of such topics could be done, thus stimulating much further research and eventually the development of policy of action.*

The choosing of sites to investigate follow the same path as the choice of a single case study, but with in the addition of geographical restraint. Statoil was the only organisation with a large enough base and with people that could easily be reached for meetings and interviews. They are also the organisation that are furthest along in the process of implementing the required hardware needed for the LogisticHub. Later in the research when I had to extend my sites, these were selected by who I could get in connect with, and that were interested in providing me with information (UD4,UD5).

When using single case studies it becomes even more important to have several data sources, this is to best ensure validity and reliability of the research findings. (Yin, 2009) mentions that a clear description of the research data sources and how these contributed to the findings are important in order ensure the reliability and validity of the findings. I have therefore presented the different data collected and used in this research in Chapter 3, and attempted to reference the data to findings as the paper has been written. This to make it possible for other researchers to trace my findings back to the data source.

Chapter 5 represents the initial data gathered. These were data gathered from several documents and meetings, resulting in an analysis that turned out to to become the different models presented in this chapter. This is not completely in line with the case study method, but because of delays in the process of developing the LogistcHub, I had to go ahead and start parts of the process on my own. I would have liked to have gotten more direct feedback on the models in this early stage, as for example when investigating the system development techniques, but when the time came to get this feedback, an official document on the LogistcHub was released, containing much of the feedback I sought. It also became hard getting feedback from the stakeholders as their focus changed at this point. I think the research would have benefited from feedback on my charts and diagrams done in Chapter 5.

A good deal of data collected came from meetings. In these meetings I was presented with a lot of data from second hand sources. This was information obtained by stakeholders at the meeting, where they had discussed the different issues with other stakeholders. The data collected from these meetings can be hard to reproduce, since these meetings were not part of the data collecting in the original research design, they were not recorded and only short informal notes were made. In addition, some of these meetings were with my external counsellor, this might lower the reliability of the study, but due to the need to redesign the research and change the research question, I had to use every data source I had available to complete this research. If there had been more time available, I could have gotten more of the information from the meetings confirmed by interviews directly with the different sources that initially stated them, and thus increasing the reliability of the study.

When I had to redesign the research, I tried to get more information from other sources and sites. This became difficult because many of the stakeholders that were interesting to interview, were not very knowledgeable about the subject of the LogisticHub. They knew it was going to be developed and that is was supposed to be a trace and tracking system. Additionally, when I contacted people they referred me to people I had already talked to. Some of the interviews had to be done by telephone. This was not optimal, but had to be done this way due to geographical restrictions that made it hard to make direct contact with some of these sources (IN2,IN3,ME2,ME8).

Due to time restraints and delays in the development process of the LogisticHub, I had to simulate some of the development process on my own. This was done in part of the original research design, and that made this data not fit completely with the new research design, but because I felt this to be important data that revealed interesting things about the research, I chose to present this as part of the data collected. This might break with the recommendations of case studies, since the researcher is not supposed to actively engage in this manner. Thus, as the researcher, I risk bringing my own bias opinions to the table, lowering the validity of the research.

Chapter 9

Conclusion

With this research I wanted to examine the process of developing a semantic hub and compare it to the process of developing other information systems. The semantic hub in this case needed to facilitate a large number of different stakeholders and their organisations. These organisation have different views on the semantic hub, as to how they desire the outcome to be, and how what benefits are to be gain from it. This makes it important that all types of stakeholders are represented as part of the development. The techniques used in the development of other information systems worked well in the development of the semantic hub. It might require the use of certain techniques, but these techniques could easily be used to cross check and validate the system with the development techniques.

Semantic Web technologies are relatively new and untested in an enterprise environment making it necessary to test certain technologies before one commits to using them. This is most likely not a specific thing to the process of developing a semantic hub, but more to the process of developing an information system utilising new technologies.

9.1 Future Research

This research has focused on finding the differences in the development process of a semantic hub compared to the development process of other information systems. The research has yielded some interesting findings that need to be confirmed by future research. I therefore would like to present two future research possibilities to validate these research findings.

The first is to continue to follow the development process of the LogisticHub, because this development process is still not completed. A continued research on this development process to examine the actual development techniques used in the next stages of the development, and an examination into the technologies that they end up using to implement the LogisticHub. To see if these also match the development process of other information systems.

The second path of future research could be to look at other cases and cross check, these to investigate whether or not my findings are valid in other cases This might be difficult to conduct, due to the lack of semantic hubs that are built in enterprise environments.This is only the start of semantic hubs. I think that in the future, a lot more of these types of system will be developed, making it important for us developers to understand the subtle differences from the development of semantic hubs compared to other information systems.

Appendix A

Z-015 checklist form

Lagre og avslutt   Skriv ut sjekkliste

Skriv ut interface hook-up data sheet

Checklist for containers and equipment

Norsok Standard Z-015

Statoil

Units ID no: HTCC 188	Receiver contact pers.: Rekvirent
Vendor/Eier: Furmanite AS	Vendor contact pers.: Nils Kristian Vik
Platform/Innretning: 1 - Åsgard B	Approved for use in zone: Non-hazardous

Forhåndsgodkjenning ansvarlig person i selskapet, ref kapittel 4.8.3:

Telefon:

Navn: Dato godkjenning mottatt:

Lag ny logg her --> Du kan ikke lage ny logg

U98 - Annet ikke Ex-sikkert utstyr på ramme

OK = Approved NA = Not Applicable NC = Not Checked D = Defective or missing	Ref. Z-015	STATUS leverandør	STATUS mottaker	KOMMENTARER
1-MEKANISK				
1.1 Merking og skilting	4.6	OK	OK	
1.2 Årlig kontroll (løfteutstyr, sertifikat)-**lev. må legge inn siste kontrolldato**	4.4.1	19.12.2012	NA	Note 1.2
1.3 Mekanisk tilstand (løfteører/punkter, struktur, etc)	4.4.1	OK	OK	
3-ELEKTRO				
3.1 Kabler, nipler, utstyr etc.	4.4.6.1	OK	OK	
3.2 Isolasjonsmotstand (megging)-**lev. må legge inn siste kontrolldato**	4.8.1	19.12.2012	OK	
3.3 Jording (NB bonding av struktur)	4.4.6.3	OK	OK	
3.4 All installasjon utvendig i Ex-utførelse	4.5.1	NA	NA	Note 3.4
4-OPPKOBLING (GRENSESNITT) MOT PLATTFORM				
4.1 Kraftforsyning (spenning, frekvens og støpsler)	4.7	OK	OK	
4.4 Alarmsignal til kontrollrom med støpsel	4.7	NA	NA	Note 4.4
6-DOKUMENTASJON				
6.1 Løftesertifikat	4.8	NA	NA	Note 6.1
6.2 Øvrig teknisk dokumentasjon (tegninger, datablad etc)	4.8	OK	NC	
6.3 Sertifikat for utstyr med spesielle krav (EEx...x, PSV)	4.8	NA	NC	Note 6.3
6.4 Vedlikeholdsprogram	4.8	OK	NC	
6.5 Vedlikeholdsjournal	4.8	OK	NC	
6.6 Brukermanual	4.8	OK	NC	

Generelle kommentarer:

All dokumntasjon ligger på data og som operatør tar med ut på ÅSGB ved ankomst plattform.
Kun visuell EX-sjekk av TT Statoil SRMN.

Leverandør signatur - kyndig løfteteknisk/mekanisk	Nils Kristian Vik	19.12.2012
Leverandør signatur - elektro fagmann	Nils Kristian Vik	19.12.2012
Karakter på kontainer/utstyr: ... Karakter ved mottak fra leverandør: 1 ... Karakter ved forsendelse: 2 Mottaker signatur (ansvarlig for utyllling av sjekkliste)	Per Einar Ersvik	07.01.2013

Send mail til kontaktperson ombord for midlertidig utstyr

Bibliography

Aduna. (2012). Sesame. Retrieved November 12, 2013, from http://www.openrdf.org

Alani, H., Kalfoglou, Y., O'Hara, K. & Shadbolt, N. (2005). Towards a killer app for the semantic web. *The Semantic Web ISWC ...* 829–843. Retrieved from http://www.springerlink.com/index/u65h72676h117170.pdf

Avison, D. & Fitzgerald, G. (2006). *Information systems development: methodologies, techniques & tools* (McGraw-Hill, Ed.). McGraw-Hill. Retrieved from http://books.google.com/books?id=xk9dPwAACAAJ%5C&pgis=1

Becker, B., Dawson, P., Devine, K., Hannum, C., Hill, S., Leydens, J., ... Palmquist, M. (2012). Case Studies. Writing@CSU. Retrieved March 12, 2012, from http://writing.colostate.edu/guides/guide.cfm?guideid=60

Beckett, D. [Dave] & MacBride, B. (2004). RDF/XML Syntax Specification (Revised). Retrieved February 22, 2012, from http://www.w3.org/TR/REC-rdf-syntax/

Beckett, D. [David] & Berners-Lee, T. (2008). Turtle—Terse RDF Triple Language. Retrieved from http://www.dajobe.org/2004/01/turtle/

Belchior, M., Schwabe, D. & Parreiras, F. (2012). Role-Based access control for model-driven web applications. *Web Engineering*, 106–120. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-31753-8%5C_8

Benbasat, I., Goldstein, D. K. & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly*, *11*(3), pages. Retrieved from http://www.jstor.org/stable/248684

Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The Semantic Web. *Scientific American*. Lecture Notes in Computer Science, *284*(5), 34–43. doi:10.1038/scientificamerican0501-34

Carroll, J. J., Bizer, C., Hayes, P. & Stickler, P. (2005a, December). Named graphs. *Web Semantics: Science, Services and Agents on the World Wide Web*, *3*(4), 247–267. doi:10.1016/j.websem.2005.09.001

Carroll, J. J., Bizer, C., Hayes, P. & Stickler, P. (2005b). Named graphs, provenance and trust. *Proceedings of the 14th international conference on World Wide Web - WWW '05*, 613. doi:10.1145/1060745.1060835

Costabello, L. & Villata, S. (2012). SHI3LD: an access control framework for the mobile web of data. *Proceedings of the 23rd ...* 311–312. Retrieved from http://dl.acm.org/citation.cfm?id=2310049

Costabello, L., Villata, S., Delaforge, N. & Gandon, F. (2012). Linked data access goes mobile: Context-aware authorization for graph stores. *... on Linked Data on ... 2012*. Retrieved from http://hal.archives-ouvertes.fr/hal-00691256/

Decker, S., Melnik, S., Harmelen, F. V., Fensel, D., Klein, M., Broekstra, J., ... Horrocks, I. (2000). The Semantic Web: the roles of XML and RDF. *IEEE Internet Computing*, *4*(5), 63–73. doi:10.1109/4236.877487

E&P Information Management Association. (2012, March). About EPIM. Retrieved March 19, 2013, from http://www.epim.no/epim/menu/home

Feigenbaum, L., Williams, G. T., Clark, K. G. & Torres, E. (2013, March). SPARQL 1.1 Protocol. Retrieved May 30, 2013, from http://www.w3.org/TR/sparql11-protocol/

Flouris, G., Fundulaki, I., Michou, M. & Antoniou, G. (2010). Controlling access to RDF graphs. *Future Internet-FIS 2010*, 107–117. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-15877-3%5C_12

Fowler, M. (2004). *UML distilled*. Retrieved from http://books.google.com/books?hl=en%5C&lr=%5C&id=nHZslSr1gJAC%5C&oi=fnd%5C&pg=PP19%5C&dq=UML+distilled%5C&ots=V77

Gabillon, A. & Letouzey, L. (2010, September). A View Based Access Control Model for SPARQL. *2010 Fourth International Conference on Network and System Security*, 105–112. doi:10.1109/NSS.2010.35

Gable, G., Sedera, D. & Chan, T. (2008). Re-conceptualizing Information System Success : the IS-Impact Measurement Model. *... of the Association for Information ... 9*, 377–408. Retrieved from http://eprints.qut.edu.au/14058

Garlik, S. H., Seaborne, A. & Prud'hommeaux, E. (2012). SPARQL 1.1 Query Language. Retrieved from http://www.w3.org/TR/sparql11-query/

Gruninger, M. & Fox, M. S. (1995). Methodology for the Design and Evaluation of Ontologies. *Industrial Engineering*, *95*, 1–10. Retrieved from http://ibict.phlnet.com.br/anexos/grninger95methodology.pdf

Hayes, P. (2004). RDF Semantics. W3C. Retrieved from http://www.w3.org/TR/rdf-mt/

Hebeler, J., Fisher, M., Blace, R. & Perez-Lopez, A. (2011). *Semantic Web Programming*. John Wiley & Sons. Retrieved from http://books.google.com/books?id=rtcIFmXkysEC%5C&pgis=1

Hendler, J. (2001). Agents and the Semantic Web. *IEEE Intelligent Systems*, *16*(2), 30–37. doi:10.1109/5254.920597

Jung, J.-G., Park, S.-B., Cha, S.-J. & Jo, G.-S. (2008). Semantic Web Service for Freight Management System. *2008 IEEE International Workshop on Semantic Computing and Applications*, 132–138. doi:10.1109/IWSCA.2008.27

Kaplan, B. & Duchon, D. (1988). Combining Qualitative and Quantitative Methods in Information Systems Research: A Case Study. *MIS Quarterly*, *12*(4), 571–586. doi:10.2307/249133

Kapoor, B. & Sharma, S. (2010). A Comparative Study Ontology Building Tools for Semantic Web Applications. *International Journal*, *1*(July), 1–13. doi:10.5121/ijwest.2010.1301

Kim, J., Jung, K. & Park, S. (2008). An introduction to authorization conflict problem in RDF access control. *Knowledge-Based Intelligent Information and ...* 583–592. Retrieved from http://www.springerlink.com/index/W761727H586R4168.pdf

Kinsella, S., Bojars, U., Harth, A., Breslin, J. G. & Decker, S. (2008, July). An Interactive Map of Semantic Web Ontology Usage. *2008 12th International Conference Information Visualisation*, 179–184. doi:10.1109/IV.2008.60

Kogut, P., Cranefield, S., Hart, L., Dutra, M., Baclawski, K., Kokar, M. & Smith, J. (2002). UML for Ontology Development Why Use UML to Develop Ontologies ? Retrieved from http://www.vistology.com/papers/KER3.PDF

Kvale, S. & Brinkmann, S. (2009). *InterViews: Learning the Craft of Qualitative Research Interviewing* (V. Knight, S. Connelly, L. Habib, A. Virding & J. Robinson, Eds.). Interviews: Learning the Craft of Qualitative Research Interviewing. Sage Publications. Retrieved from http://www.amazon.co.uk/dp/0761925422

Lassila, O., Swick, R. R., Wide, W. & Consortium, W. (1999). Resource Description Framework (RDF) Model and Syntax Specification. http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/#intro. Retrieved February 28, 2012, from http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/%5C#intro

Norsk Standard. (2004). NORSOK STANDARD Z015: Temporary equipment.

Norsk Standard. (2013). Z-015 Temporary equipment: Additional information. Retrieved May 3, 2013, from http://www.standard.no/en/Sectors/Petroleum/NORSOK-Standard-Categories/Z-Temporary-Equipm/Z-0152/

Norwegian Oil and Gas Association. (2010). *OLF GUIDELINE No.112 DEPLOYMENT OF RADIO FREQUENCY IDENTIFICATION (RFID) IN THE OIL AND GAS INDUSTRY*. Retrieved from http://www.norskoljeoggass.no/en/Publica/Guidelines/Integrerte-operasjonerIntegrated-operations/112/

Noy, N. F. & Mcguinness, D. L. (2000). Ontology Development 101 : A Guide to Creating Your First Ontology. *Development*. Lecture Notes in Computer Science, *32*(1), 1–25. doi:10.1016/j.artmed.2004.01.014

Object Management Group. (2011). BUSINESS PROCESS MODEL AND NOTATION (BPMN). (January). Retrieved from http://www.omg.org/spec/BPMN/2.0/PDF

Object Management Group. (2013). What is UML. Retrieved May 17, 2013, from http://www.omg.org/gettingstarted/what%5C_is%5C_uml.htm

Park, S. (2008, April). An RDF triple based access authorization specification considering subsumption relationships. *2008 IEEE 24th International Conference on Data Engineering Workshop*, 462–469. doi:10.1109/ICDEW.2008.4498361

Pérez, J., Arenas, M. & Gutierrez, C. (2006). Semantics and Complexity of SPARQL. *The Semantic Web-ISWC 2006*, 30–43. Retrieved from http://www.springerlink.com/index/p14356kpx6631057.pdf

Petter, S., DeLone, W. & McLean, E. (2008, June). Measuring information systems success: models, dimensions, measures, and interrelationships. *European Journal of Information Systems*, *17*(3), 236–263. doi:10.1057/ejis.2008.15

Protégé. (2013). Protégé. Retrieved January 5, 2013, from http://protege.stanford.edu

Silverman, D. (2006). Credible Qualitative Research BT - Interpreting Qualitative Data 3rd Edition. In *Interpreting qualitative data 3rd edition* (8, pp. 271–314). SAGE Publications.

The Norwegian Oil and Gas Association. (2012, November). Logistichub Requirements Specification.

Thi, P., Thuy, T., Lee, Y.-k., Lee, S. & Jeong, B.-s. (2007). Transforming Valid XML Documents into RDF via RDF Schema. *Knowledge Creation Diffusion Utilization*, 35–40. doi:10.1109/NWESP.2007.23

Uschold, M. (1996). Building Ontologies: Towards a Unified Methodology. *16th Annual Conf of the British Computer Society Specialist Group on Expert Systems.* Technical Report of the Artificial Intelligence Applications Institute, (September), 16–18. doi:10.1.1.39.9075

Uschold, M. & King, M. (1995). Towards a Methodology for Building Ontologies. In D. Skuce (Ed.), *Methodology* (Vol. 80, *July*, pp. 275–280). Citeseer. doi:10.1.1.55.5357

Veltman, K. H. (2001, January). Syntactic and semantic interoperability: New approaches to knowledge and the semantic web. *New Review of Information Networking*, *7*(1), 159–183. doi:10.1080/13614570109516975

Villata, S., Delaforge, N. & Gandon, F. (2011). S4AC Vocabulary Specification 0.2. Retrieved January 10, 2013, from http://ns.inria.fr/s4ac/v2/s4ac%5C_v2.html

Watkins, E. & Nicole, D. (2006). Named graphs as a mechanism for reasoning about provenance. *Frontiers of WWW Research and Development- ...* 943–948. Retrieved from http://www.springerlink.com/index/71117k1j2x16463q.pdf

Yin, R. K. (2009). *Case Study Research: Design and Methods* (L. Bickman & D. J. Rog, Eds.). Applied Social Research Methods Series. Sage Publications. doi:10.1097/FCH.0b013e31822dda9e