Modeling RDF Data for MetOcean Information Systems

Kamaluddeen Usman Danyaro and Jafreezal Jaafar
Department of Computer & Information Sciences, Universiti Teknologi PETRONAS, Tronoh Perak, Malaysia
Email: kamaluddeen_g01325@utp.com.my, jafreez@petronas.com.my

M. S. Liew
Department of Civil Engineering, Universiti Teknologi PETRONAS, Tronoh Perak, Malaysia
Email: shahir_liew@petronas.com.my

Abstract—**This paper suggests a Resource Description Framework (RDF) model for effective handling of distributed data relating to the Semantic Web. RDF is a data model that integrates aptly structured data. The World Wide Web is facing great challenges in data retrieval as a result of increase and request of knowledge from different sources. This brings the issue of information overload. Therefore, in minimizing this challenge, the paper presents the process of modeling RDF knowledge base which is the essential step of Semantic Web development. As a state-of-the-art research, a meteorological and oceanographic dataset has been used to provide the basic concepts of RDF serialization. The result accessed through the developed SPARQL endpoint. Thus, it implies that the approach is sufficient for querying as well as data representation of distributed data.**

Index Terms—**Resource Description Framework (RDF), RDF Model, Semantic Web, MetOcean**

I. Introduction

Since from the inception of Resource Description Framework (RDF) in 1990s, various processes have been followed as a way to produce the ideal model or structure for language specification. In 2004, there were extended RDF specifications for representing semantic data that have been standardized by World Wide Web Consortium (W3C). RDF is the most essential source of Semantic Web that works with various things on the networks. It manages and handles the distributed data which gives techniques for data representation [1]. The representation of data from unstructured information into organized information are called modeling [1]. Nevertheless, RDF is a language for specifying languages. It is among the three modeling languages (RDF, RDFs and Web Ontology Language) of Semantic Web. These languages stipulate the expressivity and structure within the Web knowledge. Semantic Web is about information systems integration in meaningful way. Therefore, it depends on information sharing between community as well as the systems.

An RDF application follows the typical sources of data. This can be done by integrating the data with relational database management system and later incorporate it with the Web through Uniform Resource Identifiers (URIs). URI also offers a good level of data presentation and enabling shared data from different sources; global references. The infrastructure is laid over the distributed networks through Uniform Resource Locator (URL). Then become interoperable by exchanging and handling the eXtensible Markup Language (XML) data [2, 3]. XML has been the mediator between human and computers (on the Web) with the help of standard syntax that produce documents. Although, XML and database supply the consistency of every Web that initiate all sorts of data models [1]. However, XML as being a data model alone cannot handle large amount of data and system interoperability. It is because the XML provides syntax not semantics for data representation over the web. Nevertheless, many Web database systems are not on Linked Data model structure. For instance, the database of Meteorological and Oceanographic (MetOcean) information system is built on XML and XML metadata only [4].

MetOcean is a meteorological and oceanographic industry that handles large amounts of data and metadata. A lot of companies and research industries depend on its data that has been built on XML schema. However, these tight on the capabilities of describing data or metadata in meaningful which leads to the problems of information overload. Information overload is the process of getting information that might not be sufficiently organized as a result of rapid advancement of information and communication technology [5-7].

The aim of this paper is to define and describe the distributed data of MetOcean in a meaningful way. In this regard, we design a network graph, perhaps a triplestore that stores graphs of MetOcean's semantic repository and able to be queried as linked data.

The structure of this paper has the following outline. Section 2 provides the RDF and its basic concepts. Section 3 discusses the need for RDF model in meteorological and oceanographic information systems. Section 4 provides the related works. In section 5 we present the proposed RDF data model consisting of the structure of MetOcean's triplestore, RDF representation

doi:10.4304/jcp.9.2.432-440

and graphical data representation. Section 6 contains conclusions and the way forward for the study.

II. Related Work

Allemang and Hendler [1] used Turtle as their RDF serialization for better compactness of the model. They display the model from N-Triple and the qnames. Also, N-Triples are employed to express different data models of RDF [8], where it can be applied for testing and debugging the dataset applications.

For the blank nodes, the documents require RDF serialization [9]. When querying the data relationship, Oren et al. [10] have evaluated ActiveRDF that manipulates data using bnode in both quantitative and qualitative sense. Therefore, bnodes are helping nodes that explicitly used in RDF which requires no URI. They cannot be addressed or referenced globally with respect to URI [1, 11].

RDF Serialization is a foundation for RDF data model interpretation. Zander and Schandl [12] serialized RDF data on mobile device in which the files for the devices have been utilized for RDF standardization formats. However, [12] also found the delay in processing RDF graphs as a result of loaded working memory (RAM) before querying the SPARQL Protocol and RDF Query Language (SPARQL). While Yoo [13], proposes three methods using the SPARQL query for querying the knowledge base of Semantic Web. He provides the inference engine that interprets the meaning of information on RDF knowledge base. He found that the machine could be assessed by individual user requirement using reasonable query method. The finding was in hybrid query processing of hotel search. Unlike Rodriguez [14], that used serialization language to translate the Ripple query. In serializing RDF graphs, the run-time along with the number of triples stored have proven to be linear with regards to his framework [12].

An entity relational model was developed for the better expressiveness of what has been called semi-structured data models which include RDF and microformats [15]. The query expressiveness of RDF databases has sufficient power in answering the complex graph-shaped queries where mostly are designed to provide flexibility to users [15-17]. This implies the uniqueness of Delbru et al. [15] findings that the system provides efficient index maintenance and faster query processing.

Thus, almost all the above authors find that their systems sufficient for query or indexing large amount of data, which is also among the targets of this paper. Furthermore, due to the expressive nature of Turtle, we follow [1] and (Becket and Berners-Lee [18]) to have the strong binding between the global and local (URIs and qnames) for our RDF development.

III. The Need for RDF Model in MetOcean IS

There were many researches on various models regarding sea or particularly South East Asia Fine Grid Hindcast (SEAFINE). These researches [19-23] targeted prediction, simulation, and hindcast analysis. Obviously, the primary target is to retrieve data in an accurate and easier manner. However, these models have not yet been implemented the RDF or Linked Data accessibility. Consequently, this paper is based upon an RDF development for Semantic Web. The SEAFINE includes the MetOcean's participants namely: BP, ChevronTexaco, ConocoPhillips, Murphy Oil, Statoil, Total, BP and Sarawak Shell Berhad [24-26].

Oceanweather Inc. and Joint Industry Project (JIP) created SEAMOS (South East Asia Meteorological and Oceanographic Hindcast Study) hindcast in 1992 with intention to investigate the storm influences basically in southern South China Sea [20, 26]. The specific objectives for SEAFINE is to provide: a fine grid wind and wave hindcast that constitutes 50-year continuous data, large grid resolution for waves and wind fields, and 20-50 years fine resolutions for current hindcast [24].

Notwithstanding, the fundamental reason why SEAMOS has been produced is because of the need of higher spatial resolution models by SEAFINE which itself is a hindcast model. As laid out in [26], the resolution is located within the areas like Gulf of Thailand, Java Sea and Makassar Strait. Moreover, in SEAFINE alone, the MetOcean data comprises: (i) wind and wave data series with grid of 6km x 6km, a period of 50 years data (1956-2007) and (ii) ocean current series with grid of 12km x 12km, period of 20 years data (1981-2007) [26].

Thus, hindcasting or hindcast models are viewed as common tools for data specification and testing in the meteorological and oceanographic environment numerically. Users are able to retrieve the data in a single query if only they depend on such tools. This is because the current MetOcean system models are not interoperable. Thus, the RDF is considered as the heart of Semantic Web data application which will provide data integration and solve the problem of interoperability.

IV. RDF and the Information Exchange

In this part, we provide and elucidate the basic concepts of the RDF data model as well as the RDF serialization. These concepts consist of publishing RDF data and representing it in text format so that it could be understood easily. RDF syntax modeling is the fundamental root of structuring data in a Semantic Web environment.

The RDF graphs enable individuals to share data with other people via the W3C standardization. The expression or specification of these data by means of triples using a language statement can be referred as resource description framework or RDF [11, 19]. RDF reduces ambiguities by conceptualizing the complete resources of anything. The resources are connected through URI. URI is the superset of Uniform Resource Locator (URL) that describes how digital information retrieves data over RDF resources [11]. As we described in [27], Figure 1 describes the RDF structure which consists of triple or subject, predicate and object (s, p, o).

Subject	Predicate	Object
UTP	partOf	PETRONAS
Total	isLocatedIn	"North Bali"
MetOcean	owned	SEAFINE
SEAFINE	hasCompany	Shell
Kinabalu	hasFrequency	“20 minutes”

Figure 1. Triple relationship.

A. The Linked Data

The effective use of Web standard as well as common data models leverages the power of Web applications to publish the data on the global data space. The establishment is always machine-processable. These give the avenue of establishing the data globally and make it to be the interconnected data space [14, 28, 29]. The connection is established with the help of application programming interfaces (APIs). A database is connected through agents (such as JDBC) then to Web server and finally arrives to user. The user receives the documents (XML) through HTTP or HTML via a browser. The XML data is delivered through many kinds of connections before it reached to a destination. Therefore, to publish data as structured data on the Web, Semantic technologies pull the connections. It is simply by translating the data into RDF and Web Ontology Language (OWL) and then links the data from one source to another. This also establishes the knowledge structure of the database.

The connection of interrelated datasets on the Web is called linked data [28, 30, 31]. The things or resources on the Web could be implicit in XML data. These however after connection with linked data components will then be the explicit links that could easily be accessible by the users [28]. The main linked data components are RDF, RDF-Schema (RDFs), Web Ontology Language (OWL) and other publish linked data APIs. Listing the triples inform of RDF representation is can be referred as RDF serialization.

B. RDF/XML

RDF/XML is considered as the most common serialization formats [8]. In other words, it is a method of basic XML serialization of RDF. Web interfaces present information to the user in HTML or/and XML documents. The Web data depend on the URL processes. In Semantic Web, the qualified URI processes are based on RDF/XML which contains qname declaratives. See the examples of the qname declaratives; *rdf* and *spd* below:

```
<rdf : RDF
  xmlns : spd
="http://www.metoceansemweb.com/seafine/participant#"
   xmlns : spd =http://www.w3.org/1999/02/22-rdf-syntax-ns#>
     <spd : Particpant
       rdf : about = http://www. metoceansemweb.com
       /seafine/participant#Participant1>
           <spd : ID>1</spd : ID>
           <spd : Company>Shell</spd : Company>
           <spd : Location>M1 Field West</spd : Location>
           <spd : SamplingPeriod>1993-1994</spd : SamplingPeriod>
    </spd : Particpant>
    <spd : Particpant
       rdf : about = http://www. metoceansemweb.com
       /seafine/participant#Participant2>
           <spd : Location>Satun Platform</spd : Location>
           <spd : SamplingPeriod>1998</spd : SamplingPeriod>
           <spd : ID>2</spd : ID>
           <spd : Company>Chevron</spd : Company>
    </spd : Particpant>
    <spd : Particpant
       rdf : about = http://www. metoceansemweb.com
       /seafine/participant#Participant3>
           <spd : Location>Kinabalu</spd : Location>
           <spd : ID>3</spd : ID>
           <spd : Company>Shell</spd : Company>
           <spd : SamplingPeriod>1992-1994</spd : SamplingPeriod>
    </spd : Particpant>
<rdf : RDF>
        -     -    -
```

C. Turtle

Turtle or Terse RDF Triple Language is serialization format that can be applied to express data in an RDF data model. It is a superset of N-Triples which is compatible with N-Triples and N3 [18].

The turtle starts with local qnames that leads the expression of triples by setting subject, predicate, and object in order, as in the following:

```
spd : Participant1  rdf : type  spd : Participant  .
```

Representing the Turtle in triple format ends with a period (.). Similarly, when representing many triples a semicolon (;) is used at the end of the line which signifies the presence of another triple. With respect to Table 1, the Turtle representation of triples can be listed as:

```
spd : Participant1   rdf : type  spd : Participant ;
        spd : Participant_ID “1”  ;
        spd : Participant_Location “M1 Field West”  ;
        spd : Participant_Company “Shell”  ;
        spd : Participant_Sampling_period“1993-1994”  .

spd : Participant2   rdf : type   spd : Participant ;
        spd : Participant_ID “2”  ;
        spd : Participant_Location “Satun Platform”  ;
        spd : Participant_Company “Chevron”  ;
        spd : Participant_ Sampling_period  “1998”  .

 spd : Participant3   rdf: type  spd : Participant ;
        spd : Participant_ID “3”  ;
        spd : Participant_Location “Kinabalu”  ;
        spd : Participant_Company “Shell”  ;
        spd : Participant_Sampling_period“1992-1994”  .

spd : Participant4   rdf : type  spd : Participant ;
        spd : Participant_ID “4”  ;
        spd : Participant_Location “Jerneh”  ;
        spd : Participant_Company “Exxon-Mobil”  ;
        spd : Participant_Sampling_period“1995-1996”  .

spd : Participant5   rdf : type  spd : Participant ;
        spd : Participant_ID “5”  ;
        spd : Participant_Location “Kikeh Field”  ;
        spd : Participant_Company “Murphy”  ;
        spd : Participant_Sampling_period“2003-2005”  .

spd : Participant6   rdf: type  spd : Participant ;
        spd : Participant_ID “6”  ;
        spd : Participant_Location “Tinggi”  ;
        spd : Participant_Company “Exxon-Mobil”  ;
        spd : Participant_Sampling_period“1995-1996”  .

spd : Participant7   rdf : type  spd : Participant ;
        spd : Participant_ID “7”  ;
        spd : Participant_Location “Kebabangan”  ;
        spd : Participant_Company “Shell”  ;
        spd : Participant_Sampling_period“1994-1999”  .

spd : Participant8   rdf : type  spd : Participant ;
        spd : Participant_ID “8”  ;
        spd : Participant_Location “Semangkok”  ;
        spd : Participant_Company “Exxon-Mobil”  ;
        spd : Participant_Sampling_periodn “1994-1996”  .

spd : Participant9   rdf: type  spd : Participant ;
        spd : Participant_ID “9”  ;
        spd : Participant_Location “Doyong”  ;
        spd : Participant_Company “Exxon-Mobil”  ;
        spd : Participant_Sampling_period “1995-1996”  .

spd : Participant10   rdf : type  spd : Participant ;
        spd : Participant_ID “10”  ;
        spd : Participant_Location “Fairley-Baram”  ;
        spd : Participant_Company “Shell”  ;
        spd : Participant_Sampling_period “1987-1988”  .

spd : Participant11   rdf : type  spd : Participant ;
        spd : Participant_ID “11”  ;
        spd : Participant_Location “Donggala”  ;
        spd : Participant_Company “Total”  ;
        spd : Participant_Sampling_period“2002-2003”  .

spd : Participant12   rdf: type  spd : Participant ;
        spd : Participant_ID “12”  ;
        spd : Participant_Location “B11 Field”  ;
        spd : Participant_Company “Shell”  ;
        spd : Participant_Sampling_period “1994-1995”  .

spd : Participant13   rdf : type  spd : Participant ;
        spd : Participant_ID “13”  ;
        spd : Participant_Location “Baronia Field”  ;
        spd : Participant_Company “Shell”  ;
        spd : Participant_Sampling_period “1993-1994”  .

spd : Participant14   rdf : type  spd : Participant ;
        spd : Participant_ID “14” ;
        spd : Participant_Location “Gendalo”  ;
        spd : Participant_Company “Chevron”  ;
        spd : Participant_Sampling_period “2005-2006”  .

spd : Participant15   rdf: type  spd : Participant ;
        spd : Participant_ID “15” ;
        spd : Participant_Location “Bukittua”  ;
        spd : Participant_Company “Conoco”  ;
        spd : Participant_Sampling_period “2006”  .

spd : Participant16   rdf : type  spd : Participant ;
        spd : Participant_ID “16” ;
        spd : Participant_Location “Baram” ;
        spd : Participant_Company “Shell”  ;
        spd : Participant_Sampling_period “1992-1993”  .

spd : Participant17   rdf : type  spd : Participant ;
        spd : Participant_ID “17” ;
        spd : Participant_Location “North Bali”  ;
        spd : Participant_Company “Total”  ;
        spd : Participant_Sampling_period “2004”  .
```

D. N-Triples

This is a serialization format which allows the machine to be mapped and process data in a more precise recorded manner [32]. N-triple has high properties in becoming machine-readable than the other serialization formats. It is also a full line-based URIs. See the examples of URIs below:

```
<http://www. metoceansemweb.com /seafine/participant#
  http:// metoceansemweb.com /seafine/participant#participant1
  http:// metoceansemweb.com /seafine/participant#participant
```

This cannot easily be presented in this paper because of its line-based property. However, we replace http://www.metoceansemweb.com/seafine/participant# with “spd : participant : participant” as a shorter form of URI the we have:

```
<spd : participant  : participant1> <rdf : type : participant _ID>  “1” .
<spd : participant : participant1> <rdf : type : participant _location> “M1 Field West ” .
<spd  :  participant  t:  participant1>   <rdf  :  type  :  participant _company> <shell> .
```

Therefore, each URI in N-Triples must be in a single line that starts with “<” and end with “>”. N-Triples can also be represented in a basic RDF/XML format. It is a simple and plain text encoding format for RDF graph which makes it a good choice for streaming data applications [33, 34].

E. N3

N3 or Notation 3 is made for easy expression of triples between the backward and forward links [35]. In other words, N3 is the text-encoded and abstract compact RDF syntax that allows the extension of the query expressiveness [1, 14, 15, 36]. Nevertheless, it helps in shorting the URI. For instance, a URI for *kinabalu*; <metoceansemweb.com/seafine/participant/kinabalu> can be reduced to *metocean:kinabalu*. This provides entity identification as well as prefix declaration which usually occurs at the beginning of documents.

F. Blank Node

In a situation where the URIs' resources could be referenced or identified, the blank (anonymous) nodes would be used to address the subjects and objects using URIs [11]. RDF provides the connection of nodes without any URI which is then called bnode or blank nodes. This usually occurs when modeling many-valued relationships [11].

V. METOCEAN RDF DATA MODEL

In this section, we show the RDF model that can be developed through designing triplestore. We design a query for subject, predicate and object. These triples represent the index relating to their permutations within the program.

Therefore, we maintained the standard querying representation pattern (s, p, o). For instance, in Figure 2 the intuitive directed graph describes the knowledge representation of a simple fact. This triple interpretation means that the subject *Kinabalu* is related to the object *20 minutes* (of sampling frequency) by the predicate *hasFrequency*. In other words, the predicate *hasFrequency* is connecting the *Kinabalu* with *20 minutes* value (connecting resources together).

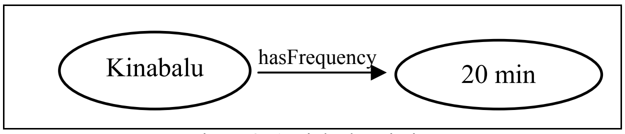

Figure 2. A triple description

A. MetOcean TripleStore Structure

The Semantic Web itself cannot provide ideal representation devoid of the utilization of any data structure of the system. The Semantic representation enables metadata applications to connect with RDF and XML for better operation. Specifically, the connection of the MetOcean Information Systems (IS) in the non-RDF MetOcean database of Figure 3 is the bedrock of the architectural model. It comprises the data and metadata which have been preserved in MetOcean IS. The MetOcean XML is the XML that would be serialized in the triplestore. This allows the heterogeneity of the data sources. The RDF semantic application is placed in order to bridge the gap between the semantic application and the raw data. Nonetheless, these three components have direct connectivity with the triplestore which causes the repository to keep MetOcean's knowledge. Most of the research industries and various oil and gas sectors would depend on this knowledge base. At this juncture, the data representation could only be through RDF. However, the SPARQL client is the endpoint that the user will retrieve the query through.

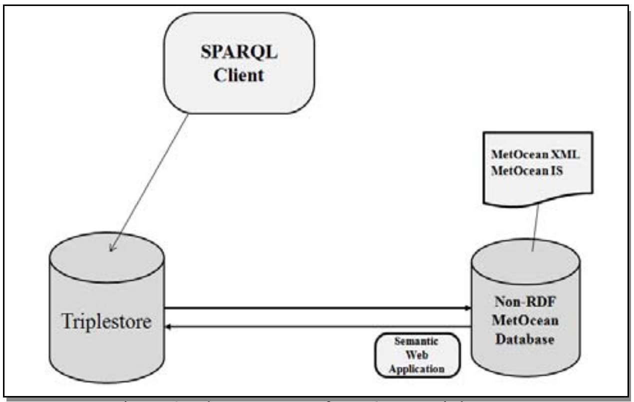

Figure 3. The structure of MetOcean triplestore

B. RDF Representation

In this work, we consider the data available by SEAFINE participants. We refer the reader to Table I of [27] that shows the basic metadata and locations for current measurements. First off, we converted the datasets into tabular format for simplification. Then we produced the sample data as triples. It describes the observational dataset that includes: location, company, water depth, sampling frequency, sampling duration and sampling period.

To represent the data in RDF while the column and row as the information and single entity respectively. If a row has distinct entity then it most have single URI [1]. In Semantic Web, we only need a global unique identifier in order to have the single URI for each identifier in the database.

Now, let *spd* be the namespace, then we concatenate the name of the table (Participant) with a unique key that expresses the identifier. Impliedly, *spd:participant1*, *spd:participant2*, *spd:participant3*, and so on correspond to the location, company, sampling frequency, sampling duration, water depth and sampling period properties in RDF representation.

To have the global unique identifiers for these properties, we merged the column name and table name. Then property become: *spd-:-participant_ID*, *spd_:_participant_Location*, *spd_:_participant_Company*, etc. Therefore, each cell in the table has one triple which expresses the table information. See Table I which consists of 21 triples. However, with respect to Figure 5, we have 119 triples that consist of subjects, predicates and objects. The subjects and predicates are in RDF formats while objects are literal formats which include integers and strings.

TABLE I.
PARTICIPANT'S TRIPLE REPRESENTATION

SUBJECT	PREDICATE	OBJECT
spd:Participant1	spd:Participant_ID	1
spd:Participant1	spd:Participant_Location	M1 Field West
spd:Participant1	spd:Participant_Company	Shell
spd:Participant1	spd:Participant_Sampling_Frequency	20 minutes
spd:Participant1	spd:Participant_Sampling_Duration	16 months
spd:Participant1	spd:Participant_Water_Depth	139m
spd:Participant1	spd:Participant_Sampling_Period	1993-1994
spd:Participant2	spd:Participant_ID	2
spd:Participant2	spd:Participant_Location	Satun Platform
spd:Participant2	spd:Participant_Company	Chevron
spd:Participant2	spd:Participant_Sampling_Frequency	20 minutes
spd:Participant2	spd:Participant_Sampling_Duration	12 months
spd:Participant2	spd:Participant_Water_Depth	71m
spd:Participant2	spd:Participant_Sampling_Period	1998
spd:Participant3	spd:Participant_ID	3
spd:Participant3	spd:Participant_Location	Kinabalu
spd:Participant3	spd:Participant_Company	Shell
spd:Participant3	spd:Participant_Sampling_Frequency	20 minutes
spd:Participant3	spd:Participant_Sampling_Duration	25 months
spd:Participant3	spd:Participant_Water_Depth	61m
spd:Participant3	spd:Participant_Sampling_Period	1992-1994
-	-	-

Clearly, each row in Table I corresponds to a participant. If we add one triple per row, we could represent it in RDF. This also describes the type of individual in a row which finally results the RDF graphs.

As described in section I and II, MetOcean has a huge amount of data. Therefore, it needs a lot of RDF networks for linking its relational data, see Figure 5. Nevertheless, the model provides the linking statement graphically; the portion of participant1 in Figure 4 depicts the SEAFINE participant1 statement. We also derive the directed graphs from the statements. Apparently, this would be the required data model that can handle huge amount of data and metadata.

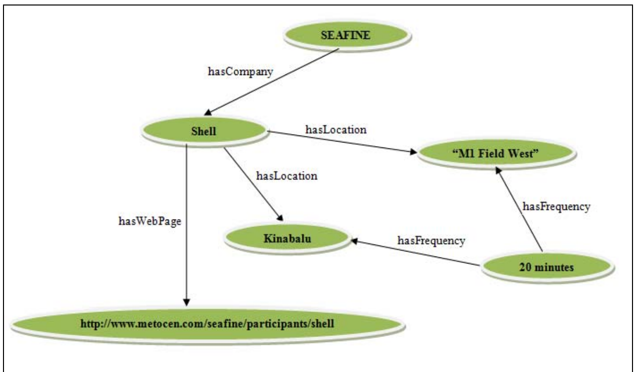

Figure 4. Linking graph

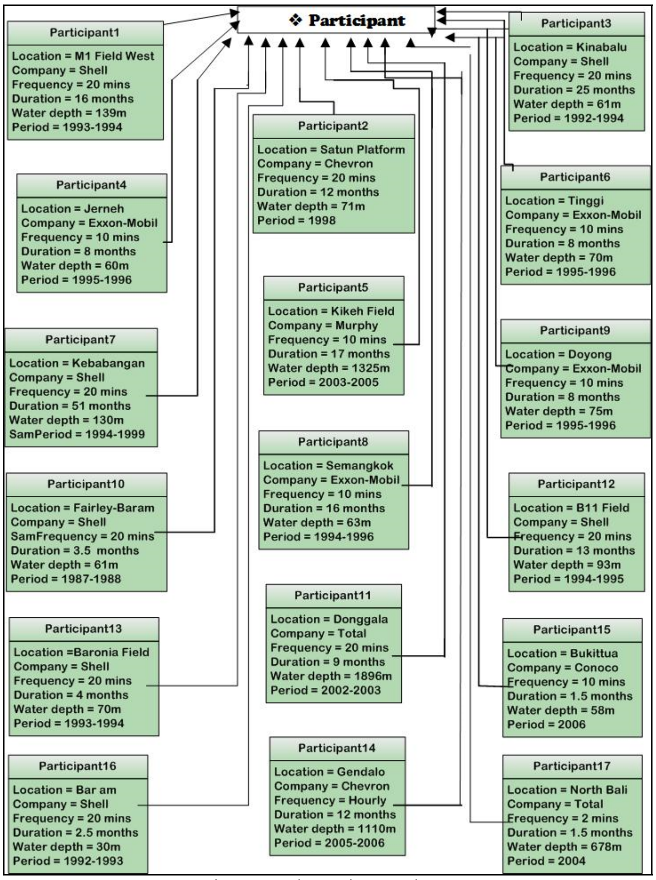

Figure 5. The entity graph

The model is visually represented and simplified. In this sense, the triplestore is established and the data can be queried using SPARQL, See Figure 7 for a simple query of participant1.

C. MetOcean and the Distributed Data

This work focuses on making the data globally distributed on the Web. Therefore, exchanging the data between the distributed users that are relevant to meteorological and oceanographic environment can be achieved through the developing the knowledge base.

The system has been designed to make the namespace and the URI for MetOcean IS knowledge usable. It then followed by developing the model knowledge using the class description. The knowledge description of instance identifies the classes with URIrefs. See the domain definition below:

```
<rdf:RDF
xmlns="http://www.metoceansemweb.com/participants.owl#"
   xml:base="http://www.metoceansemweb.com/participants.owl"
   xmlns:participants2="&participants;20"
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
   xmlns:owl="http://www.w3.org/2002/07/owl#"
   xmlns:xsd=http://www.w3.org/2001/XMLSchema#

xmlns:participant="&participants;www.metoceansemweb.com/seafi
ne/participant/"
   xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#

xmlns:participants="http://www.metoceansemweb.com/participant
s.owl#">
          - - -
    <owl:DatatypeProperty rdf:about="&participants;hasLocation">
      <rdf:type rdf:resource="&owl;FunctionalProperty"/>
      <rdfs:domain rdf:resource="&participants;Shell"/>
      <rdfs:range rdf:resource="&xsd;string"/>
    </owl:DatatypeProperty>
          - - -
</rdf:RDF>
```

VI. Result

With respect to the discussions above, we provide the SPARQL endpoint in which a user can query the result from. The implementation was done using D2R server[1]. D2R is an application for publishing RDF data from database to Semantic Web. The serialization was conducted based on N3 notation. The dataset used in this study was a meteorological data.

The result appears based on user's selection. Suppose a user wants to count the categories of participants, he can formulate the aggregate function as:

```
PREFIX spd: <http://www.metoceansemweb.com/>

SELECT (Count(DISTINCT ?Participant) AS ?PETRONAS)

WHERE { ?Item rdf:type ?Participant }

LIMIT 10
```

This allows smooth connections between datasets. It selects the match *rdf:type* participant and presents the result as the count. See Figure 7, the SPARQL endpoint.

We queried the concepts from the developed SPARQL endpoint. Apparently it can be discovered that MetOcean data can be mapped and create rich pathways by means of querying the resources. The accuracy of the result has been evaluated using F-measure. See Table II. The F-measure is the well-known method used in information retrieval [37, 38]. It consists of precision and recall as the two metrics, with non-negative value, α. Consider $N_{correct}$ to be the number of correct or relevant ratio and $N_{incorrect}$ to be the number of incorrect or irrelevant ratio.

$$F_{\alpha} = \frac{(1+\alpha)\cdot \text{precision} \cdot \text{recall}}{\alpha \cdot \text{precision} + \text{recall}}$$

$$\text{Precision} = \frac{N_{correct}}{N_{correct} + N_{incorrect}}$$

$$\text{Recall} = \frac{N_{correct}}{N_{total}}$$

TABLE II.
CONCEPT RETRIEVAL

Name	Records contained	Relevance	Recall	Precision
Participants	1092	752	0.6886	0.8624
SeaPlaces	1715	901	0.5254	0.9037
Location	976	680	0.6967	0.9577

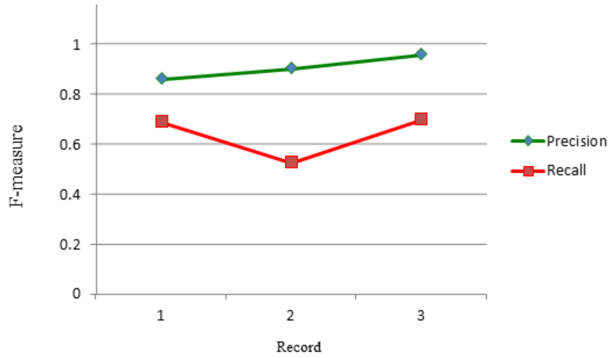

Figure 6: Graph query comparison

Now, applying this method in our experiment, it signifies the effectiveness of the system. Therefore, this indexing pattern is amenable for larger database and more sophisticated queries. Nevertheless, it implies the expressiveness of RDF data available in the MetOcean database.

Conclusion

One of the main fundamental roots of building Semantic Web is RDF database. Through this work, we found that the RDF model is sufficient for building the triplestore in the meteorological and oceanographic sector. The data are set in a manner to which each triple could be queried using SPARQL. The result suggests that it can be employed to query a lot of graphs. Moreover, due to the need of Semantic Web in meteorological environment, this system is quite flexible to users. It provides better approach for Semantic Web development.

In the future, we intend to develop more sophisticated queries as a way for developing Semantic Web in the meteorological and oceanographic environment. Also, we will then infer the system using fuzzy logic systems for rule-based inference reasoning with RDF as well as OWL.

[1] http://d2rq.org/

ACKNOWLEDGEMENT

This work was partially supported by Universiti Teknologi PETRONAS (UTP) and PETRONAS Carigali Sdn Bhd. We thank reviewers for their helpful comments.

Figure 7: SPARQL endpoint

REFERENCES

[1] D. Allemang and J. Hendler, *Semantic Web for the working ontologist: effective modeling in RDFS and OWL*: Morgan Kaufmann, 2011.

[2] S. Powers, *Practical RDF*: O'Reilly \& Associates, Inc., 2003.

[3] A. Swartz. (2003, January 3, 2012). *application/rdf+xml Media Type Registration*. Available: http://www.w3.org/2001/sw/RDFCore/mediatype-registration

[4] Metocean. (17 May 2010). Available: http://www.metocean.co.nz/cms

[5] A. Edmunds and A. Morris, "The problem of information overload in business organisations: a review of the literature," *International Journal of Information Management,* vol. 20, pp. 17-28, 2// 2000.

[6] S. R. Hiltz and M. Turoff, "Structuring computer-mediated communication systems to avoid information overload," *Commun. ACM,* vol. 28, pp. 680-689, 1985.

[7] V. Sugumaran and J. A. Gulla, *Applied semantic web technologies*. USA: CRC Press Taylor & Franceis Group, 2012.

[8] T. Segaran, C. Evans, and J. Taylor, *Programming the semantic web*: O'Reilly Media, 2009.

[9] K. U. Danyaro, J. Jaafar, and M. S. Liew, "Tractability method for ontology development of Semantic Web," in *Semantic Technology and Information Retrieval (STAIR), 2011 International Conference on*, 2011, pp. 34-39.

[10] E. Oren, B. Heitmann, and S. Decker, "ActiveRDF: Embedding Semantic Web data into object-oriented languages," *Web Semant.,* vol. 6, pp. 191-202, 2008.

[11] P. Hitzler, S. Rudolph, and M. Krötzsch, *Foundations of semantic web technologies*: Chapman & Hall/CRC, 2009.

[12] S. Zander and B. Schandl, "Context-driven RDF data replication on mobile devices," *Semantic Web,* vol. 3, pp. 131-155, 2012.

[13] D. Yoo, "Hybrid query processing for personalized information retrieval on the Semantic Web," *Knowledge-Based Systems,* vol. 27, pp. 211-218, 3// 2012.

[14] M. A. Rodriguez, "The RDF virtual machine," *Knowledge-Based Systems,* vol. 24, pp. 890-903, 2011.

[15] R. Delbru, S. Campinas, and G. Tummarello, "Searching web data: An entity retrieval and high-performance indexing model," *Web Semantics: Science, Services and Agents on the World Wide Web,* vol. 10, pp. 33-58, 1// 2012.

[16] S. Scheglmann, G. Groener, S. Staab, and R. Lämmel, "Incompleteness-aware Programming with RDF Data," 2013.

[17] H. MahmoudiNasab and S. Sakr, "An Experimental Evaluation of Relational RDF Storage and Querying Techniques," in *Database Systems for Advanced Applications*. vol. 6193, M. Yoshikawa, X. Meng, T. Yumoto, Q. Ma, L. Sun, and C. Watanabe, Eds., ed Berlin Heidelberg: Springer, 2010, pp. 215-226.

[18] D. Becket and T. Berners-Lee. (2011, 20 January 2013). *Turtle-Terse RDF Triple Language. W3C Team Submission 28 March 2011*. Available: http://www.w3.org/TeamSubmission/turtle/

[19] K. Eik, "Iceberg drift modelling and validation of applied metocean hindcast data," *Cold Regions Science and Technology,* vol. 57, pp. 67-90, 7// 2009.
[20] Z. Mayeetae, M. S. Liew, and K. V. John, "Parametric study on environmetal loads of hindcast and measured full scale data," in *National Postgraduate Conference (NPC), 2011*, 2011, pp. 1-6.
[21] J. Lawrence, H. Kofoed-Hansen, and C. Chevalier, "High-resolution metocean modelling at EMEC's (UK) marine energy test sites," in *Proc. of the 8th European Wave and Tidal Energy Conference*, 2009.
[22] A. Woolf, B. Lawrence, R. Lowry, K. K. van Dam, R. Cramer, M. Gutierrez*, et al.*, "Climate Science Modelling Language: standards-based markup for metocean data," in *Proceedings of 85th meeting of American Meteorological Society*, 2005.
[23] M. A. Hashim, J. Jaafar, and S. M. Taib, "Interval type 2 fuzzy logic for significant wave height forecasting," in *Intelligent Systems Design and Applications (ISDA), 2012 12th International Conference on*, 2012, pp. 794-798.
[24] (3 January). *SEAFINE (SEAMOS-SOUTH FINE GRID HINDCAST)*. Available: http://info.ogp.org.uk/metocean/JIPs/SEAFINE.asp
[25] S. Spall and R. Rayner, "SEAFINE 3D currents Validation of HYCOM 3D Current Modelling of the South China Sea" unpublished.
[26] I. Oceanweather, "Seamos-South Fine Grid Hindcast (SEAFINE): A Joint Industry Project Final Report" unpublished.
[27] K. U. Danyaro, J. Jaafar, and M. S. Liew, "An RDF model for meteorological and oceanographic information systems," in *Computer & Information Science (ICCIS), 2012 International Conference on*, 2012, pp. 480-484.
[28] T. Heath and C. Bizer, "Linked data: Evolving the web into a global data space," *Synthesis lectures on the semantic web: theory and technology,* vol. 1, pp. 1-136, 2011.
[29] M. Hert, G. Reif, S. Marsella, and H. C. Gall, "UpLink: a Linked Data editor for RDB-to-RDF data," presented at the Proceedings of the 7th International Conference on Semantic Systems, Graz, Austria, 2011.
[30] W3C. (2013, 12 December 2013). *Linked Data*. Available: http://www.w3.org/standards/semanticweb/data
[31] C. Bizer, T. Heath, and T. Berners-Lee, "Linked Data - The Story So Far," *International Journal on Semantic Web and Information Systems (IJSWIS),* vol. 5, pp. 1-22, 2009.
[32] D. Beckett and B. McBride. (2004, 20 Jan 2013). *RDF/XML syntax specification (revised)*. Available: http://www.w3.org/TR/REC-rdf-syntax/
[33] D. Beckett. (2013, 16 October 2013). *N-Triples : A line-based syntax for an RDF graph*. Available: http://www.w3.org/TR/n-triples/
[34] J. Hebeler, M. Fisher, R. Blace, and A. Perez-Lopez, *Semantic web programming*: John Wiley & Sons, 2011.
[35] T. Berners-lee, D. Connolly, L. Kagal, Y. Scharf, and J. Hendler, "N3logic: A logical framework for the world wide web," *Theory Pract. Log. Program.,* vol. 8, pp. 249-269, 2008.
[36] A. Hogan, A. Harth, and A. Polleres, "Scalable Authoritative OWL Reasoning for the Web," ed: IGI Global, 2009, pp. 49-90.
[37] Y. Sasaki, "The truth of the F-measure," *Teach Tutor mater,* pp. 1-5, 2007.
[38] M. Ehrig and S. Staab, "QOM – Quick Ontology Mapping," in *The Semantic Web – ISWC 2004*. vol. 3298, S. McIlraith, D. Plexousakis, and F. Harmelen, Eds., ed: Springer Berlin Heidelberg, 2004, pp. 683-697.

Kamaluddeen Usman Danyaro is currently a PhD student and graduate assistant at the department of computer and information sciences, Universiti Teknologi PETRONAS, Malaysia. He obtained his MSc. in Business IT from Northumbria University, UK. His BSc. degree was in Mathematical Science at Bayero University, Kano. His research focuses on the development of Semantic Web in Meteorological and Oceanographic sector, as well as computational methods using fuzzy logic for Semantic Web development.

Jafreezal Jaafar is the Head of Computer and Information Sciences Department, Universiti Teknologi PETRONAS (UTP), Malaysia. He obtains B.Sc in Computer Scienc from Universiti Teknologi Malaysia in 1998, MAppSc. (IT) from RMIT University (Australia) in 2002 and PhD in AI from University of Edinburgh (Scotland, UK) in 2009. Previously he has worked as a System Engineer for several years. He has been Assessor for Malaysia Qualification Agencies (MQA) for Computer Science since 2009. At the University level, he is among the Senate members, panel for curriculum Review, Internal Quality Auditor and QAT team for ISO implementation. His main research interest is Agent Systems and Soft Computing. He also had secured research grants and actively produced numerous journals, conferences and workshop papers.

M. S. Liew is the Head of the offshore Engineering Center and Civil Engineering Department at the Universiti Teknologi PETRONAS. He obtained his PhD from Texas Tech University, USA. He is currently researching on MetOcean conditions on Malaysian water and is actively engaged in the transformation of oil and gas sector for the local companies.