Market Requirements Document

Feature Name: Semantic Web Support

Version: 2   **Date Submitted:** 01/18/2008   **Completed By:** Leon Guzenda
Version: 1   **Date Submitted:** 11/09/2007   **Completed By:** Leon Guzenda

Description of the Problem

Background

"The **Semantic Web** is an evolving extension of the World Wide Web in which web content can be expressed not only in natural language, but also in a format that can be read and used by software agents, thus permitting them to find, share and integrate information more easily.

Currently, the World Wide Web is based mainly on documents written in Hypertext Markup Language (HTML), a markup convention that is used for coding a body of text interspersed with multimedia objects such as images and interactive forms. Metadata tags provide a method by which computers can read the content of web pages. For example:

```
 <meta name="keywords" content="computing, computer studies, computer">
<meta name="description" content="xxxx... ">
<meta name="author" content="xxxx">
```

The semantic web takes the concept further. It involves publishing the data in a language, Resource Description Framework (RDF), specifically for data, so that it can be manipulated and combined just as can data files on a local computer.

The HTML language describes documents and the links between them. RDF, by contrast, describes arbitrary things such as people, meetings, or airplane parts. ." – Wikipedia

Example (from Wikipedia)

In the English language statement *'New York has the postal abbreviation NY'* , *'New York'* would be the subject, *'has the postal abbreviation'* the predicate and *'NY'* the object.

Encoded as an RDF triple, the subject and predicate would have to be resources named by URIs. The object could be a resource or literal element. For example, in the Notation_3 ("N-Triples") form of RDF, the statement might look like:

```
        <urn:states:New%20York> <http://purl.org/dc/terms/alternative> "NY" .
```
In this example: "urn:states:New%20York" is the URI for a resource that denotes the U.S. state New York; "http://purl.org/dc/terms/alternative" is the URI for a predicate; and "NY" is a literal string.

Note that the URIs chosen here are not standard, and don't need to be, as long as their meaning is known to whatever is reading them.

N-Triples is just one of several standard serialization formats for RDF. The triple above can also be equivalently represented in the standard RDF/XML format as:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:terms="http://purl.org/dc/terms/">
     <rdf:Description rdf:about="urn:x-states:New%20York">
          <terms:alternative>NY</terms:alternative>
     </rdf:Description>
</rdf:RDF>
```

Problems

1. Objectivity/DB does not support the schema languages used by the various components of the semantic web stack.
2. Neither is it particularly efficient at storing and manipulating constructs that are primarily closely related text strings.
3. The internal schema representation is designed for relatively fixed structure languages, such as C++ and Java, rather than self-defining languages, such as Generalized Markup of Defined Objects (GDMO) and those used within the World Wide Web Consortium (W3C).

Impact

Customers wishing to store data from and for the Semantic Web will have to write their own metadata and data handling interfaces to Objectivity/DB, probably using Objectivity for Java. This is likely to be a considerable barrier to the adoption of Objectivity/DB in this rapidly evolving marketplace.

Description of the Requested Feature

The goal is to provide a high performance, scalable, distributed (grid-enabled), reliable database environment for Semantic Web applications and repositories. This will involve providing, at least:

a) Support for RDF Schema and RDF instance manipulation.

b) Support for XML Schema and other XML standards, such as XPath and XQuery.

c) Support for the Web Ontology Language (OWL).

d) Support for SPARQL, the RDF query language.

Features a) and b) could be implemented independently, but features c) and d) depend on the other two features. So, this MRD is a high level index to three other MRDs covering RDF, XML and OWL, respectively.

Part of an existing feature or does it require another feature, if so, which one?

This feature will add at least one and perhaps three new, optional, Objectivity products.

How is this problem being solved now, and why isn't that acceptable?

Customers who need to add support for the Semantic Web to existing products based on Objectivity/DB will have to build their own metadata and data handling tools. Prospects are more likely to look elsewhere for COTS solutions. There is a danger that some existing customers may also look for alternative solutions, either COTS or Open Source.

What languages must support this capability?
- Java
- C#.

Which platforms must be supported?
- Tier 1 at first and all platforms eventually.

Do any competitors already have this feature?
- [Progress Software] ObjectStore and Versant have limited XML support.
- The RDBMS vendors have full XML support.
- Oracle and IBM have a growing number of RDF/OWL tools.
- DATA-GRID is developing a fully compliant, Internet enabled OWL database for delivery in Spring 2008.
- Franz AllegroGraph is a high performance RDF Triple DBMS.

Customers who require this feature
- The Intelligence Community.
- The manufacturing, petrochemical, telecom and financial markets.
- Boeing has recently expressed interest in storing RDF and OWL in Objectivity/DB.

Revenue at risk, or which could be won
- Analysts predict that the global Semantic Web market could be worth $75 Billion annually from 2010 onwards.

When is this required?
- Before 2009, if possible.

Additional Notes

We will also need:
- Marketing collateral.
- Updated Technical Publications.

- New QA material.