Information Management Meets the Semantic Web

Mr. Salim K. Semy
The MITRE Corporation
202 Burlington Road
Bedford, MA 01730
ssemy@mitre.org

Dr. Mark Linderman
AFRL Information Directorate/IFSE
525 Brooks Road
Rome, NY 13441-4505
mark.linderman@rl.af.mil

Ms. Mary K. Pulvermacher
The MITRE Corporation
1155 Academy Park Loop
Colorado Springs, CO 80910
pulver@mitre.org

ABSTRACT

Finding the right information at the right time and in the right format becomes increasingly difficult as more information from myriad producers is made available to increasingly diverse communities of information consumers. The development of approaches to effectively manage this information and facilitate automated processing will help to address the challenges of a burgeoning information environment. Approaches to help overcome these challenges continue to emerge. This paper considers the convergence of enabling technologies from two information sharing approaches – the Joint Battlespace Infosphere (JBI) and the Semantic Web. The JBI facilitates and manages information sharing between producers and consumers, while the Semantic Web defines the semantics of the universe of web-based information. This paper examines the interplay of the JBI, as an example of an information management infrastructure, and the Semantic Web. We examine several facets of information management that will benefit from the Semantic Web as well as identify issues addressed by information management that will need to be addressed for mission-critical application of the Semantic Web. Finally, this paper discusses fundamental differences between the JBI and the Semantic Web that emanate from their current application contexts. We conclude with an overall perspective on their relationship and highlight areas of future research.

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: Online Information Services; J.7 [**Computer Applications**]: Computers in Other Systems

General Terms

Management; Design

Keywords

Semantic Web, Information Management, Joint Battlespace Infosphere, Ontology, Publish and Subscribe

1. INTRODUCTION

Information sharing has been a goal since the start of communication. As we move to a web-based world, where there exists an enormous amount of information with unlimited connections among this information, it becomes even more challenging to make available the desired information when you want it, where you want it, and in the desired format. Technologies that both provide Information Management and facilitate automation, through Machine to Machine (M2M) interaction, will help us overcome these challenges.

The focus of this paper is to compare two information sharing approaches, the Joint Battlespace Infosphere (JBI) and the Semantic Web. The JBI concept [9, 10] was developed in a United States military context and serves as a foundation to facilitate information sharing between information producers and consumers. Its goal is to provide consumers the right information at the right time, disseminated and displayed in the right format. The Semantic Web [2] defines the semantics of web-based information precisely enough so a machine can understand it.

This paper asserts that both the JBI and Semantic Web aim to provide a mechanism to, among other things, find "the needle in the haystack", even if this needle has multiple representations. Both approaches are impacted by a common set of information technology trends [1]. These trends include:

Ubiquity: There are billions of computers. "Everyone" has access to data and computational resources and many provide data.

Complexity: There are trillions of objects to track. This complexity is hidden from users through the use of services, components, and abstraction.

Interactivity: Users are online constantly. Everyone is a participant.

Globalization: Services are available globally, using global resources.

Dis-Intermediation: Movement is toward more direct producer to consumer communication, minimizing the role of the "middle man".

Agility: Finding the right resources, at the right time, to use them in often unforeseen ways. This includes the ability to evolve a capability over time without changing the underlying infrastructure.

Although the JBI and the Semantic Web are affected by similar information technology trends, they operate in different environments. These cultural differences imply a different set of

WWW 2004, May 17-22, 2004, New York, NY USA.
ACM xxx.

[1] Grasso, Al, "Information Technology in the Commercial Age", Briefing from the "Digitization and Information Superiority Workshop" hosted by General Richard B. Myers, Commander in Chief, NORAD and USSPACE and the Commander, Air Force Space Command, Peterson AFB, Colorado, January 7, 2000

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **NOV 2003**	2. REPORT TYPE	3. DATES COVERED **00-00-2003 to 00-00-2003**

4. TITLE AND SUBTITLE **Information Management Meets the Semantic Web**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **MITRE Corporation,202 Burlington Road,Bedford,MA,01730-1420**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**9**	

constraints. The Semantic Web inherits both the strengths and weaknesses of the World Wide Web. Openness and ease of access lead to ubiquity of information but means to assess quality or suitability of information is haphazard at best. In a military environment, such as that in which a JBI might be used, the quality of information is often critical, as is the control of access to information. This paper discusses how Semantic Web technologies can help within a JBI publish and subscribe paradigm. It also describes some unique characteristics of each approach that herald fundamental differences between them. The paper concludes with an overall perspective on how JBI and Semantic Web may interoperate and highlights areas for future research.

2. BACKGROUND

We begin with a brief description of each information sharing approach. We also introduce a Newspaper use case used for illustrative examples.

2.1 The Joint Battlespace Infosphere

The United States Air Force Scientific Advisory Board (SAB) created and expanded a new concept of information management to address shortcomings of Air Force systems in two reports [9, 10]. These reports defined the Joint Battlespace Infosphere (JBI) as a combat information management system to simplify the construction, evolution and responsiveness of Air Force command and control applications. To support warfighting objectives, the JBI must balance the need for security and control with the desire for flexibility and adaptability.

Built upon a publish-subscribe foundation, the JBI would deliver information from those who have it to those who need it. In this subscription service, a consumer asks for information prior to its publication and the information is delivered to the consumer as it is published. The JBI also provides a query service to retrieve previously published information. The query service is most analogous to operations generally performed upon the Semantic Web in that it searches for patterns among previously existing information.

The information object is the JBI's atomic unit of information management. The JBI is 'information-centric' in that it is more concerned with managing information than the software that produces and consumes it. Therefore, the JBI goes beyond traditional system-centric approaches prevalent today in the Department of Defense (DoD) and service-centric approaches, such as those based upon web services, to focus on the information to be shared and managed. The JBI disseminates, persists, and controls these discrete immutable quanta of information called information objects. The JBI performs dissemination, persistence and access control decisions on an information object by information object basis. Therefore, the structure and interpretation of an information object is crucial to the successful employment of a JBI. An information object has three principle components that allow it to be managed without overly constraining the information it conveys: 1) a type, 2) a payload (essentially any finite object), and 3) metadata that describes the information object.

In order for publish-subscribe to work, subscriber applications (or simply 'subscribers') must be able to describe the information in which they are interested. Subscribers do this by providing a predicate that is evaluated against the metadata of each information object to determine if it is of interest to the subscriber. Predicates in simple implementations of JBI are solely evaluated against the metadata of each information object. While the self-contained nature of the predicate evaluation greatly simplifies the 'syntactic' evaluation of the predicate, these simple JBI implementations make it difficult to exploit semantically rich content. We discuss below how semantic information can be exploited while maintaining acceptable system performance.

The JBI does not generally inspect the information object payload; it bases its management decisions on the information object type and metadata. To do this, the JBI requires that the metadata for an object of a given type adhere to a known schema. In other words, given an object type, a schema for its metadata is available to producers and consumers of that information object type and to the JBI infrastructure as well. To facilitate this, the JBI provides a 'Metadata Repository' that stores information about known information object types. At a minimum, it includes the information object type metadata schemas. Only objects whose types registered in the metadata repository can be published. Figure 1 depicts these key information object concepts graphically.

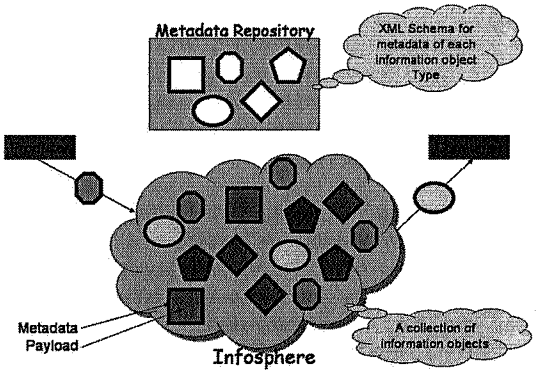
Metadata Repository
XML Schema for metadata of each information object Type
Metadata
Payload
Infosphere
A collection of information objects

Figure 1. JBI Information Objects

Based upon consumer predicates, either for subscription or query, it is the responsibility of the JBI to broker these information requests against information currently or previously published (for subscription and query, respectively). This paper examines the degree to which semantic content can be exploited to expand the sophistication of the brokering process. These enhanced brokers may expand the scope of searches by returning information objects whose relevance is determined by inference. It may also increase the precision of searches by eliminating information objects that can be shown not to be of interest. The example given in Section 3 of culling the set of trucks to those from American manufacturers based upon semantic information is an example of this.

2.2 The Semantic Web

Another approach emerging for global information management is the Semantic Web. The Semantic Web is defined as "an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."[2] A key point that bears repeating is that the Semantic Web extends the current World Wide Web incrementally. The current web makes web-based data readily accessible to humans. Using a browser, one can easily access millions of bits of data, and sift through this glut of data to find the information needed. However, most people agree that it would be highly desirable to have a machine filter this information. This requires understanding the meaning or semantics of the data. The intent of the Semantic Web is to give web-based data well-defined meaning making it machine-interpretable. Data semantics are made explicit through the use of ontologies that are then exploited by software applications.

So what is an ontology? The term ontology is not new. It originates in philosophy and has been around since the 18th century. What is relatively new is the adoption of ontologies in the web community, with the corresponding use of web technologies (web addressing, universal character set, Extensible Markup Language (XML), etc.) at their foundation. A commonly cited definition of ontology in the web community is "the specification of a conceptualization."[4] The notion is that "concepts" and their relationships to other concepts are specified precisely enough for machine interpretation. A concept may be thought of as a web resource identified by a Universal Resource Identifier (URI). Resources may either exist on the web (e.g., a document that may be retrieved) or be represented on the web (e.g., a person). An ontology captures information about these resources and the relationships between them.

Another definition from [11] is that an "ontology defines the terms used to describe and represent an area of knowledge. Ontologies are used by people, databases, and applications that need to share domain information (a domain is just a specific subject area or area of knowledge, like medicine, tool manufacturing, real estate, automobile repair, financial management, etc.). Ontologies include computer-usable definitions of basic concepts in the domain and the relationships among them." Ontologies "encode knowledge in a domain and also knowledge that spans domains. In this way, they make that knowledge reusable."

Web ontology languages are founded on a language called Resource Description Framework (RDF) and its subsequent extension called RDF Schema (RDF-S).[2] RDF represents resources as sets of triples, where each triple consists of either <Resource><Property><Resource> or <Resource><Property><PropertyValue>. These triples collectively constitute a graph. The web ontology vocabulary that is emerging as the international standard is the Web Ontology Language (OWL).[3] OWL is a semantic extension of RDF/S, providing more expressive power. On 19 August 2003, the World Wide Web Consortium (W3C) announced that OWL has become a W3C Candidate Recommendation, indicating that the specification is stable and appropriate for implementation.[4]

Ontologies provide a mechanism for machines to perform simple inferencing by combining facts together to form new facts or conclusions. For example, if one knows that B is a subclass of A and that C is a subclass of B, then one can conclude that C is a subclass of A. Another good example given in [7], and depicted graphically in Figure 2, shows that one can conclude (or infer) a new fact (i.e., Mary is the parentOf Bill) without specifically asserting that fact.

The Semantic Web brings to web-based data a formal declaration of its meaning that is precise enough to be machine-interpretable. It also provides the capability to link these descriptions both within a domain and across domains. At present a very small proportion of the data exposed on the web is marked up using Semantic Web vocabularies like RDF and OWL. As more data gets mapped to ontologies, the potential exists to achieve a "network effect."[5] To achieve the substantial participation necessary to achieve the network effect, the WWW and the Semantic Web encourage the uninhibited proliferation of content.

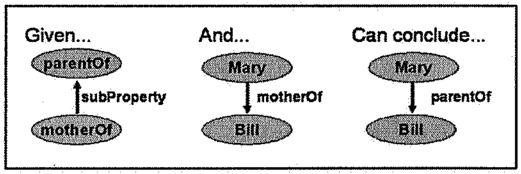

Figure 2. Simple Inferencing Example

In contrast, the JBI is an approach for managing information within domains. It assumes a less open and more controlled approach to data management. The JBI also makes a distinction between an information object's type, its metadata and its payload. The Semantic Web blurs this distinction, making it possible to represent all this information as RDF statements (i.e., sets of triples). Differences between Semantic Web and JBI approaches are discussed further in Section 4.

2.3 Use Case – Newspaper

The examples used to illustrate concepts presented in this paper focus on a Newspaper use case, specifically classified ads. A newspaper business process serves as a good example, as both publish and subscribe as well as Semantic Web technologies are applicable and may provide insight into bridging the technologies. In addition, a newspaper is a familiar concept to a large audience, and thus assumes no particular domain knowledge.

From a publish and subscribe standpoint, news articles, classified ads, etc. are represented as information objects. One can consider that a newspaper organization publishes multiple information objects, with defined relations among them, which collectively represent a newspaper publication. On the subscription side, a reader subscribes to all, or portions, of the newspaper.

From a Semantic Web standpoint, there exists a newspaper ontology, mapped to an upper concept of publication. Refer to Figure 3. Ontologies define relations between different classes of

[2] http://www.w3.org/RDF/

[3] http://www.w3.org/2001/sw/WebOnt/

[4] http://www.w3.org/2003/08/owl-pressrelease

[5] See Metcalfe's Law: "The power of the network increases exponentially by the number of computers connected to it. Therefore, every computer added to the network both uses it as a resource while adding resources in a spiral of increasing value and choice." (from: http://searchnetworking.techtarget.com/sDefinition/0,,sid7_gci214115,00.html)

publications and individual publications, as well as relationships within a publication. This paper uses classified ads as examples as ads contain semi-structured data that can be easily mapped to ontologies.

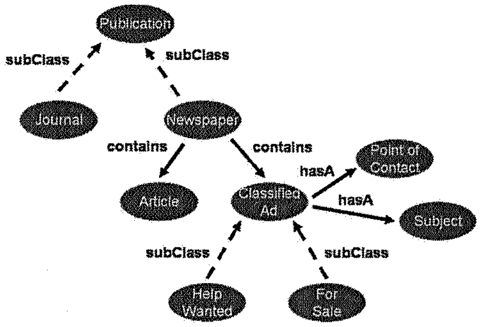

Figure 3. A Simple Publication Ontology

3. JBI PUBLISH AND SUBSCRIBE – HOW CAN THE SEMANTIC WEB HELP?

This section provides insights into how Semantic Web technologies may be used to enrich the JBI publish and subscribe approach. We discuss ideas for how Semantic Web concepts may be applied to information objects and to the process of "brokering" information objects between publishers and subscribers.

3.1 Information Objects

The concept of information object type is very similar to the notion of a class in a traditional programming sense. The metadata schema defines the properties of instances of that type. Also similar to object-oriented programming classes, information object types may be descended from other information object types by extending their metadata with additional metadata.

More generally, however, these information object types are related to one another in many ways beyond the conceptual refinement one would expect in a type or class hierarchy. Semantic Web technologies can be used to capture and exploit these relationships. Once captured, they may be used to relate information objects (instances) to one another in ways that support reasoning.

The publication ontology in Figure 3 illustrates concepts that can also be represented by information objects. A Publication information object may include metadata such as *publisher, publication date and time, format,* and *title*. A Newspaper information object, as a subclass of Publication may extend its metadata to include *local* and *price.* A Classified Ad does not extend a Newspaper object, so rather than simply adding onto the metadata of a Newspaper, it has a different set of metadata elements, such as *newspaper, item category, item description,* and *point of contact.* There may be subclasses of Classified Ad such as Help Wanted Ad and For Sale Ad that add additional metadata elements such as *duration* and *price*, respectively.

Currently, early implementations of the JBI allow information consumers to specify their information requests with predicates (or metadata constraints) over the metadata of a single information object type. In database parlance, it does not permit joins between information objects. This is certainly limiting in that the execution of a join may require that a tremendous amount of information be sent to a consumer so that it can perform the join itself. For example, to find classified ads for vehicles that have been in accidents, one would have to retrieve all the classified ads and query for accident records for each one.

Whether it is wise to support predicates that span multiple information objects is debatable. One concern is that the scale envisioned for a large JBI will almost certainly require a distributed architecture, and distributed joins, even in the far more structured relational database realm, have been shown to be very bandwidth intensive. In the less structured information space managed by a JBI, it may be very difficult, in general, to calculate the cost of evaluating such predicates and guaranteeing completeness of the result. The explicit declaration of relations, however, like in the Semantic Web can reduce the complexity of performing these joins to be (naïvely) proportional to the number of actual relations rather than the number of potential relations. Furthermore, the context that is provided by semantic content can further cull the search space.

3.2 Broker

The JBI Broker represents the minimum services required to implement the JBI Information Management approach. Such services include: allow JBI participants to connect and disconnect from the JBI, advertise and retract information publications and subscriptions, publish and unpublish information objects, and match information subscriptions with information publications, i.e. matchmaking. This section considers the use of ontologies and inferencing in the advertisement of information and the process of matchmaking.

3.2.1 Advertise Information Subscriptions and Publications

Prior to publishing or consuming an information object, participants need to make a declaration of information they will publish and/or consume. Representing these advertisements with the use of ontologies holds promise in facilitating the brokering and matchmaking process between information subscriptions and publications. A determination of how information advertisements are related to information objects that will be published, prior to the actual publication, may aid in more effective runtime matchmaking.

On the subscription side, there remains a question about the richness of a subscription, i.e. what and how should information be represented in a subscription. A part of the subscription is a set of constraints, such as temporal constraints (e.g., I'm looking for information within a particular time window) and format constraints. Mapping a date or time to a common time ontology, for example, may provide insight into the temporal association of information objects. Furthermore, a subscription may also represent usage information, such as type of information objects that are acceptable and prioritization of information objects. Representing such subtleties is an area where ontologies may be applied. Specific application of Semantic Web technologies within the subscription process is an area of future research.

3.2.2 *Matchmaking*

An essential concept in publish and subscribe-based information exchange is a process known as Matchmaking. Within an information space, represented as a collection of publications, a subscription expresses an information need. The matchmaker's role is to identify information publications that "satisfy" the information subscription. The notion of applying Semantic Web technology to address the matchmaking problem is an interesting one. Ontologies allow one to provide both structured and semantically rich metadata about information objects, expressing subtle information not expressible by traditional XML representations like an XML Schema. While an XML Schema may give syntactic and structural information about the data, the relationships between the concepts contained within the data are not articulated. An ontology makes these relationships more precise and explicit. The utility of this rich information in the matchmaking process is the focus of this section.

Prior work in applying ontologies to matchmaking algorithms serves as a good foundation and may be leveraged for information object matching within a JBI [e.g., 8, 1]. There has also been work done in semantic discovery, invocation, composition, and monitoring of Web Services. An emerging approach is the development of a web service ontology called DARPA Agent Markup Language Services (DAML-S) [6], to be replaced by OWL-S (Web Ontology Language Services). Although such efforts address diverse problems and domains, there is an analogy to matching information objects.

There are several facets of matchmaking. Matchmaking can be a multi-step process involving ever-finer specificity; it may occur at multiple levels of brokering sophistication; and there exists multiple degrees of conceptual granularity at which a match can be made (discussed in 3.2.2.4).

Multi-step of matchmaking may initially match subscriptions and publications at the class level based on common types and concepts. This step may be followed by a process of matching information objects based on their metadata values.

Multi-level matching addresses the increasing sophistication of matching from symmetric matchmaking to single ontology-based matchmaking to multiple ontology-based matchmaking. The most fundamental form is symmetric matchmaking wherein clauses within a predicate can be evaluated based solely upon explicit statements (i.e. no inference is used). It is symmetric in the sense that the form in which the predicate is constructed must exactly match the form of the data against which it is evaluated. The next level of matchmaking, single ontology-based matchmaking, uses a single ontology to provide conceptual refinement to the query. This refinement may often be inferred from semantic relationships expressed within the ontology. Finally, multiple ontology-based matchmaking leverages rules to make connections across ontologies, infer new facts and further enhance the query.

3.2.2.1 *Symmetric Matchmaking*

Symmetric matchmaking is at a rudimentary level, comparable to a traditional database query. Here, information requests contain constraints that symmetrically relate to metadata of published information. Consider an example, where a classified ad in a newspaper lists vehicles for sale. Details of the vehicle provided in the classified ad include: year, make, model, mileage, and owner information. A user may initiate a subscription such as "Notify me when a Ford Ranger appears for sale in The Boston Globe". The constraints of the subscription map directly to the metadata, of the publications; the matchmaker simply compares the subscription predicates with metadata values to determine a match.

JBI implementations that are not 'semantically-aware' use a more restrictive form of symmetric matchmaking. A JBI typically insists that predicates be well formed with respect to the metadata schema of the information objects against which they will be evaluated. In particular, this means that a predicate cannot refer to attributes that are not explicitly part of the metadata. This well-formedness restriction does not exist within the Semantic Web.

3.2.2.2 *Ontology-based Matchmaking*

A more flexible, and potentially more useful, approach is to loosely couple subscriptions to information object types such that the predicates are not bound to a single information object type. The result is that we no longer constrain the query to map solely to information object metadata and we don't need to know exactly what we want. Referring back to our truck example, a user may now say "Notify me when an American manufactured pickup appears for sale in The Boston Globe". No mention of American or pickup may appear in the classified ad. The use of an ontology, however, may provide conceptual refinement of the query such that the refined query may then be executed against the instance data.

To help illustrate this concept, a simple Vehicle Ontology is shown in Figure 4. Based on this ontology, a matchmaker is able to relate concepts, see patterns that were not immediately visible, and derive new facts. So, in our example, the ontology lets us infer 1) a pickup is a type of truck, 2) a truck is a type of vehicle, and 3) vehicles have manufacturers, such as Ford. Based on these inferences, the query may result in classified ads that contain Ford pickup trucks for sale in The Boston Globe.

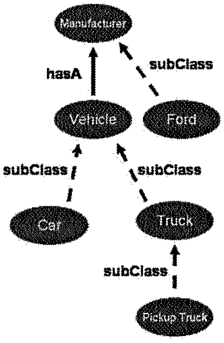

Figure 4. A Simple Vehicle Ontology

This example suggests an efficient means by which a JBI may use semantic content to service predicates that require information that is not encoded within an individual information object while still maintaining acceptable performance. Notice that most of the semantic content is relatively static, that is it does not change within our period of interest. For example, the facts that Ford is an American manufacturer and that Ford manufactures the F-150 are unchanging within the time period of interest. This observation brings to light an inherent assumption about the JBI – it is principally intended to manage information that is dynamic (i.e. that is created, used, archived, and destroyed within a relatively short period of time, such as days or weeks). Even if we cannot execute predicates over combinations of these dynamic information objects efficiently, perhaps we can efficiently perform them over a single dynamic information object and a 'knowledge-base' of static semantic content that describes the semantic context within which we operate.

3.2.2.3 Multiple Ontology-based Matchmaking

Concepts may also be related across multiple ontologies. In this case, the matchmaker makes inferences based on more than one ontology. For example, a user may query "Has Bill's Ford F-150, Vehicle Identification Number (VIN) XYZ for sale in the Boston Globe been in an accident?" As illustrated in Figure 5, a single ontology may allow one to relate an instance, such as Bill's Ford F-150, to higher-level concepts, such as the class of F-150, trucks and vehicles. This facilitates the discovery of new properties and relations, such as an F-150 is manufactured by Ford. The sample query described above, however, also requires the use of an accident ontology to provide additional pieces of the answer.

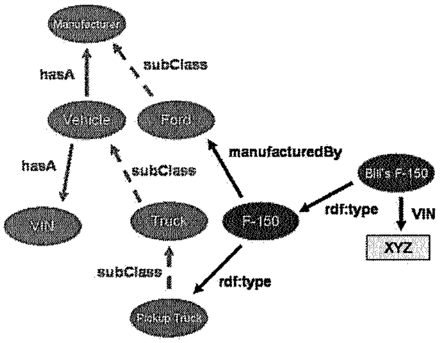

Figure 5. Instance Data Mapped to Ontology

Combining the two ontologies can help answer the query. Figure 6 depicts the power of combining ontologies. Both the Vehicle ontology and the Accident ontology refer to an instance of a Vehicle, specifically a Ford F-150 truck. Asserting that a specific F-150 with Vehicle Identification Number (VIN) XYZ in the vehicle ontology is the same as Vehicle with VIN XYZ in the accident ontology, allows further inferencing. Inference rules may allow one to make these connections across ontologies, such as equating a vehicle in the police report ontology to a vehicle in the accident ontology. For example, one rule may state "all resources in a police report that have an associated manufacturer, model, and year are Vehicles". Another rule may state that if a Vehicle in a classified ad and a Vehicle in Police Report have the same values for manufacturer, model, year, and VIN, then this is the same Vehicle. These rules allow one to infer linkages between different ontologies. Ontology translation services [3] may also be used to facilitate the discovery of concept relations across ontologies.

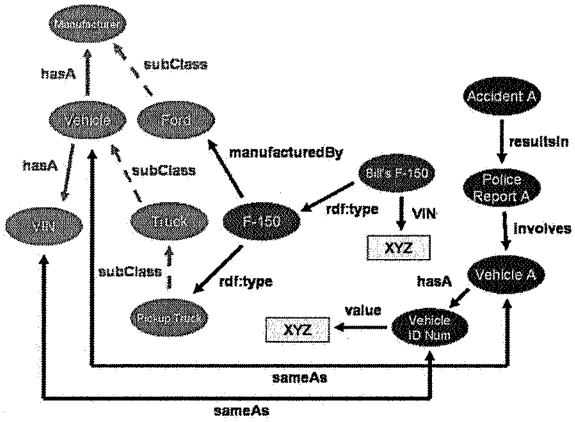

Figure 6. Multiple Ontology Mapping

3.2.2.4 Degrees of Matching

Another important aspect of a matchmaking algorithm is a definition of the term "satisfy" in the context of a publication satisfying a subscription. Information objects within a JBI may exist at different conceptual levels, and thus there are multiple degrees of matching. Utilizing ontologies to relate information objects provides the foundation to assess the likeness among information objects, based on ontological distance and relation. The categories of matching discussed below are motivated by [5]. Figure 7 provides an illustration of each type of match discussed below.

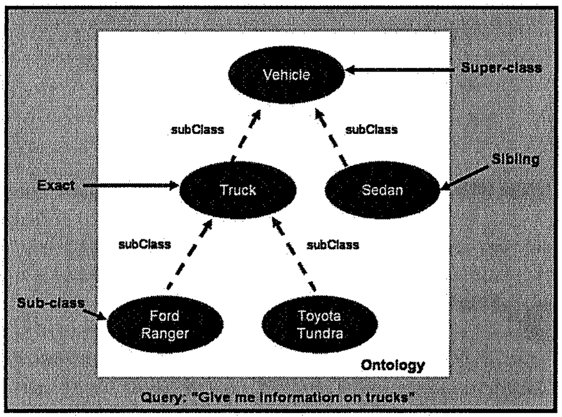

Figure 7. Degrees of Matching

Clearly, an optimal match is an exact match, where the subscription and publication are conceptually equivalent. Looking at a simple example, let's say a query asks "Give me information on trucks". The matchmaker returns an information object containing general information on trucks, such as types, manufacturer listings, etc, i.e. there is an exact match on the concept of trucks.

An exact match may not always be possible. One-to-one matches between subscriptions and publications may not exist. For example, if the query were "Give me information on trucks" and the resulting information objects contain only information on Ford pickup trucks (Ford Ranger, etc.), then obviously this does not account for all pickup trucks. By one definition, however, the query may still be considered "satisfied", as partial, but relevant, information was returned. On the other hand, suppose the query was referring to general information on pickup trucks, not particular to any models. In this case, the resulting information objects may not be very useful. This is classified as a sub-class relationship.

Information may also be returned which is conceptually at a higher level than that specified in a query, i.e. a super-class relationship. For example, a user may request information on trucks and the resulting information objects may be about vehicles. Depending on the type of information returned, there may or may not be an intersection between the subscription and resulting information objects.

Finally, there may also exist a disjoint relationship between the subscription and information objects matched, i.e. a sibling relationship. An example of this would be a user asking for information on trucks, and the resulting information objects may be about sedans. In this case, the matchmaker may have determined a relationship between pickup trucks and sedans through a common vehicle ontology. The utility of such an association in this context is questionable.

Applying Semantic Web concepts to matchmaking, as discussed, may result in varying degrees of satisfaction. While semantic markup may allow for refined matchmaking, it may inadvertently lead to large amounts of information being sent to users, some of which may not be relevant. The level and degree of matchmaking reasonable within a JBI and approaches to deal with unexpected information is an area of further research and experimentation.

4. JBI AND THE SEMANTIC WEB – AN INFORMATION MANAGEMENT ENCOUNTER

We have discussed two information sharing approaches and how Semantic Web technologies may be applied to strengthen a JBI approach. This section discusses unique characteristics of these two approaches which lead to fundamental differences between them.

As previously mentioned, the JBI concept was formulated within a U.S. military context. The mission critical application of JBI demands information management discipline and enforcement that may not be necessary in many Semantic Web applications. While arguably more critical in the JBI, the notions of content quality, access control, retraction, ownership, and information containment may also have significance for the Semantic Web, especially as it starts to be relied upon for mission critical applications, military or commercial. The following sections discuss differences between the Semantic Web and the JBI. These differences are summarized in Table 1.

Table 1. JBI and Semantic Web Comparison

Attribute	Semantic Web	JBI
Atomic Unit	RDF Triple	Information Object
Access Control	Open	Controlled
Content Quality	Unregulated	Management of publication privileges
Ownership	Voluntarily encoded or inferred from web address	Encoded with metadata and audited
Retraction	Voluntary	Enforced
Update	Replace old information	New Publication (due to information object immutability)

4.1 Atomicity and Granularity of Management

In the Semantic Web and the JBI, the granularity of management differs. In the Semantic Web, a statement (e.g. an RDF triple) is the atomic unit of semantic content. An RDF triple may contain URI references that point to locations within the document or to other documents. Exploiting semantic content may mean traversing these linkages among and between web-based resources. The concept of the web is a universe of network accessible information. Notionally, all web-based resources defined in a Semantic Web vocabulary (e.g., RDF or OWL) are part of this universe of information. This begs the question of how to handle things like cross-domain security policies when what is being passed is a set of RDF statements not a document with a prescribed format and content.

In contrast, the JBI's indivisible unit of management is an information object, which is roughly analogous to a document. These information objects are strongly managed in that they have a predefined type and each information object type has a prescribed set of metadata. This allows the creation of policy (e.g., security policy) based upon these known characteristics.

Therefore, the atomic unit in the Semantic Web and the JBI differ – they are an RDF statement and an information object, respectively. While a seemingly trivial distinction, this attribute becomes significant in discussions of policy for things like access control and security.

4.2 Access Control

Another difference between the Semantic Web and the JBI that warrants discussion is the policy area of access control. In the Semantic Web, most data is exposed on the web and therefore open to access by any user. Where access control is desired, it can currently be provided at the document level. For example, if an ontology is stored in a database management system (DBMS) then one could exploit the DBMS access control capabilities provided. However, what if one wants to access finer-grained elements like a single RDF statement? This is an active area of research. As an example, The MITRE Corporation has a research project[6] that is examining how to do policy-based sharing of fine-grained information.

In the JBI access control is strongly managed at the information object or document level. This means that a consumer can see

[6] For more information contact the Principal Investigator, A. L. Kazura at alk@mitre.org

either all or none of the information object. This policy is strongly enforced. Therefore, it may at times be desirable to separate a single information object, such as a classified ad, into two or more documents because the communities allowed access to the information are different. For example, while the content of an ad should be visible to all subscribers to the newspaper (but not necessarily everyone else), the billing information associated with the classified ad should be accessible only to the billing department of the newspaper.

With respect to access control enforcement and administration, the JBI and the Semantic Web have complementary strengths. JBI requires powerful and flexible enforcement mechanisms to control the creation and access to disparate types of information objects. Several of these mechanisms may be applicable to controlling access to semantic content for diverse communities. Semantic web ontologies may be used to infer applicable policies (e.g. access control policy) based upon relations between information object types, thus reducing administrative burden and increasing consistency.

4.3 Content Quality and Ownership

The Department of Defense is currently grappling with how to balance the need for timely information with the need for quality information. Often, the timeliest information is conflicting and additional time is required to sort through and analyze voluminous information before a consistent picture can emerge.

In the WWW, a web address provides an indication of ownership for posted data, and indirectly an indication of quality. A person is more likely to trust data from a well known and reputable company than data from "GarageBoyz.com". The Semantic Web inherits this imperfect means of assessing ownership and quality but adds the ability to voluntarily encode this information as RDF statements. However, there is no regulation of how one must encode data ownership or quality. Furthermore, as inferences are drawn based upon relations encoded across the Semantic Web, it becomes increasingly difficult to assess the quality of information. An inference engine may assert statements based upon sets of statements available within the Semantic Web, but the processing steps by which it deduced these statements may involve documents that are never made visible to the 'consumer' of the deductions. Therefore, even the imperfect (manual) means of assessing quality in the non-Semantic Web may become unobservable in the Semantic Web.

In the JBI, content quality is associated with the provider of the content. Managing information about the creator and owner of the information, as well as enforcing publication privileges, forms the basis for maintaining content quality within the JBI. Ownership is encoded in the metadata of the information object and is regulated. In addition, only certain producers are allowed to publish certain types of data, based on an assessment of their trust and reliability to provide this data. Since the source of information is an important indicator of quality, this capability supports information assessment.

Neither the JBI nor the Semantic Web provides built-in services that assess information quality. They do, however, have complementary strengths that, when used together, may go a long way towards providing assessment capabilities. One of the strengths of the JBI is its ability to capture and manage information about the creator and ownership of information objects. In a JBI context, the Semantic Web can capture the semantic significance of pedigree information that may form the basis for inferring statements about the certainty, precision, and suitability of information. Used together, pedigree-based inference and managed ownership form a strong foundation upon which to build quality assessment capabilities.

4.4 Retraction and Updates

The retraction of information is problematic in any system that maintains persistent state. Retractions and updates within a Semantic Web are totally voluntary. The owner of a URI (i.e., web address) can post what they want, update or replace it when they desire, and retract it as they see fit. This process is not controlled by anyone else. Of course, if the information producer does not want to alienate consumers of their semantically tagged data, they will want to use some best practices for managing the data (e.g., use automatic redirection for broken URLs). The point is that this is not required. Emerging Semantic Web languages do provide some mechanisms to facilitate the management of updates and retraction (e.g., OWL includes components like "owl:priorVersion", "owl:backwardCompatibleWith", and "owl:incompatibleWith"). We expect that best practices will emerge that use these mechanisms.

In a JBI, the implications of retraction and update are more apparent than in the Semantic Web because information objects are immutable. Therefore, once an information object is created, it cannot be altered, even if the producer subsequently believes that the information is no longer accurate. Furthermore, the JBI does not allow a publisher to remove an object from the information space. The JBI information management staff is responsible for removing information objects from the information space, either manually or by automatically enforced policy. There are several reasons for this, most notably that there may be other information objects that reference the original information object. In the military, it is important that it be possible to reconstruct what a person knew and the time that they knew it to ensure that warfighters have acted responsibly. To retract information as though it never existed violates this auditing requirement.

Thus, the cultural differences between the JBI and the Semantic Web are strongly evident in the area of update and retraction. However, within the JBI one could use semantic tagging to represent different types of retraction (information is incorrect, imprecise, not suitable for an intended use, out of date, etc.), each with corresponding management policies. This is clearly an area for further study.

5. FUTURE DIRECTION AND CONCLUSION

Through our examination of two information management and sharing approaches, we conclude that the use of Semantic Web technologies can add value to publish-subscribe information management approaches such as JBI. As discussed, one may use ontologies to more precisely capture relationships among information objects and use this rich information to enable inferencing, for example in the application of policy. Furthermore, ontologies may provide contextual information to enhance the matchmaking process. An information object type may map to an ontology to provide context for the published object. Leveraging single ontologies to infer relationships may allow for more generalized queries and semantics-based matchmaking. Multiple ontologies, with the addition of inference rules, may further enhance the process and permit the discovery of data and response to queries that potentially cross information object boundaries.

Our investigation has also shown that currently JBI and the Semantic Web have different priorities and cultures. In Section 4, we discussed several differences between the JBI and the Semantic Web that stem from these differing perspectives, including granularity of management, access control, content quality, ownership, retraction and updates. Presently, we believe the JBI and Semantic Web place different priorities on these characteristics due, in part, to differences in their respective operating environment. However, as the Semantic Web is applied in commercial and mission critical applications, approaches to operate within such environments must be formulated.

Looking ahead, we believe there are a number of areas that need further research. Our study has shown that Semantic Web technologies can add value to publish and subscribe information management. However, there is still a question of how best to blend semantics into a publish-subscribe approach through the use of ontologies. Specific areas of research include: identifying optimal ways to represent relations among information objects and how best to exploit these relationships, the use of ontologies to express contextual information in the subscription process, the level and degree of matchmaking reasonable within a JBI, and approaches to deal with unexpected information. Research in these areas can impact approaches for bridging the JBI with the Semantic Web.

Looking further into the future, we believe even greater challenges lie ahead. While the Semantic Web focuses primarily on the open Web (the largest data source), some applications will require tighter information management approaches. From the perspective of mission critical applications, whether they are military applications or business critical commercial applications, there is a need to enforce policy to ensure access control, content quality, timeliness and trust. The difference in atomicity (information object vs. RDF triple) has implications for security and policy, which may not be addressed in current cross-security domain solutions (i.e., security guards). Other concepts, including access control, quality, ownership, retraction and updates also need further investigation to enhance information management within the Semantic Web.

Our investigation leads us to the conclusion that there is a symbiotic relationship between the JBI and the Semantic Web. However, questions remain. We hope researchers, including those at AFRL, delve further into these topics to help solve the unanswered questions and mature the relationship between these two information sharing approaches.

6. ACKNOWLEDGMENTS

We would like to thank the following people for reviewing our paper and providing their invaluable feedback: Michael Hebert, Amy Kazura, David Kwak, Frank Manola, Leo Obrst, Scott Renner, and Arnon Rosenthal.

7. REFERENCES

[1] Ankolekar, A., M. Burstein, J. Hobbs, O. Lassila, D. Martin, S. McIlraith, S. Narayanan, M. Paolucci, T. Payne, K. Sycara, and H. Zeng. DAML-S: Semantic Markup for Web Services.

[2] Berners-Lee, T., J. Hendler, and O. Lassila. The Semantic Web. Scientific American, vol. 284, no. 5, May 2001, pp. 34–43.

[3] D. Dou, D. McDermott, and P. Qi. Ontology Translation on the Semantic Web.

[4] Gruber, T. R., 1993. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition 5: 199-200.

[5] Li, L. and I. Horrocks. A Software Framework For Matchmaking Based on Semantic Web Technology.

[6] M. Paolucci et al. Semantic Matching of Web Services Capabilities. Proceedings of the International Semantic Web Conference (ISWC02), 2002.

[7] Pease, Adam. Why Use OWL? Available online at: http://www.xfront.com/why-use-owl.html

[8] Tangmunarunkit, H., S. Decker and C. Kesselman. Ontology-based Resource Matching in the Grid – The Grid meets the Semantic Web.

[9] United States Scientific Advisory Board Report on Building the Joint Battlespace Infosphere, Volume 1: Summary, SAB-TR-99-02, December 17, 1999.

[10] United States Scientific Advisory Board Report on Information Management to Support the Warrior, SAB-TR-98-02, December 1998.

[11] W3C Candidate Recommendation. OWL Web Ontology Language Use Cases and Requirements, 18 August 2003. Available online at: http://www.w3.org/TR/webont-req/.