18th ICCRTS

Title of Paper: Making Semantic Information Work Effectively for Degraded Environments

Primary Topic: Architectures, Technologies, and Tools

Alternatives: Data, Information and Knowledge;
Collaboration, Shared Awareness, and Decision Making

Authors

Jason Bryant Matthew Paulini
Air Force Research Laboratory
RISA - Operational Info Management
525 Brooks Road
Rome, NY 13440

Point of Contact: Jason.Bryant@rl.af.mil 315-264-3259

Organization:
Air Force Research Laboratory - RISA
Operational Information Management Branch
525 Brooks Road
Rome, NY 13441

Organization Phone: 315-330-2373

Organization E-mail: Jason.Bryant@rl.af.mil

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUN 2013		**00-00-2013 to 00-00-2013**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Making Semantic Information Work Effectively for Degraded Environments	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) **Air Force Research Laboratory,RISA - Operational Info Management,525 Brooks Road,Rome,NY,13440**	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
Presented at the 18th International Command & Control Research & Technology Symposium (ICCRTS) held 19-21 June, 2013 in Alexandria, VA.

14. ABSTRACT
The challenges of effectively managing semantic technologies over disadvantaged or degraded environments are numerous and complex. One of the greatest challenges is the size of raw data. Large messages prevent semantics from being operationally effective by overflowing the available bandwidth and memory resources of degraded environments. Our approach mitigates this challenge by performing data reduction through the adoption of format recognition technologies, semantic data extractions, and the application of mission-based and role-based filters. The other challenge is that semantics are not especially effective in degraded environments due to a lack of interoperability with standardized DOD messaging formats and resource limitations for processing, storage, reasoning, and bandwidth. Our approach increases the utility of semantics by extracting attributes that correlate with DOD messaging standards and collaborate with missionbased domain ontologies, as well as to relocate the intensive processing, storage, bandwidth, and semantic extraction applications to enterprise services. This enables semantics to be utilized effectively, regardless of environment, while simultaneously informing quality of service and rolebased policy decisions.

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**	**Same as Report (SAR)**	**12**	

Abstract

The challenges of effectively managing semantic technologies over disadvantaged or degraded environments are numerous and complex. One of the greatest challenges is the size of raw data. Large messages prevent semantics from being operationally effective by overflowing the available bandwidth and memory resources of degraded environments. Our approach mitigates this challenge by performing data reduction through the adoption of format recognition technologies, semantic data extractions, and the application of mission-based and role-based filters. The other challenge is that semantics are not especially effective in degraded environments due to a lack of interoperability with standardized DOD messaging formats and resource limitations for processing, storage, reasoning, and bandwidth. Our approach increases the utility of semantics by extracting attributes that correlate with DOD messaging standards and collaborate with mission-based domain ontologies, as well as to relocate the intensive processing, storage, bandwidth, and semantic extraction applications to enterprise services. This enables semantics to be utilized effectively, regardless of environment, while simultaneously informing quality of service and role-based policy decisions.

INTRODUCTION

Disadvantaged and degraded environments can reference a variety of environment attributes, each with distinct challenges. To an end user it is inconsequential whether their applications are disadvantaged due to technological constraints such as opportunistic networking, minimal bandwidth, low-power devices, or other resource restrictions, active interference from hostile forces, or passive interference from physical features of the terrain. The need to be flexible, adaptive, effective, and highly available, regardless of constraints, is crucial. Many use cases for degraded environments deal with tactical tasks or missions, meaning that downtime and hindrance to performance is not an option. Our work shows that through the use of semantic technologies it is possible to effectively and efficiently provide mission-critical, time-sensitive information to consumers within disadvantaged and degraded environments. However, merely using semantic technologies is not the complete solution.

Semantic technologies make reuse of information possible, provide context through relationships, and allow for inferencing and analysis of the information for fusion or decision-making. These features do not come without a cost. Semantics is a specialized area of standards and research which offers a unique combination of capabilities for structured data linking, interoperability, and reasoning, yet lacks standardized best practices for a wide variety of use cases. Redundancy, persistence, ontology selection and optimization, and bulky technology formats can make it unrealistic and impractical to implement semantic applications for degraded environments. The lack of existing best practices for semantic use cases is understandable when considering the relative immaturity of Semantic Web technologies and standards. The disruptive innovation lifecycle curve of semantics implies that reconciling use cases with appropriate prescribed technological solutions are both time and resource consuming. Our work attempts to alleviate some of these pitfalls so that semantics may be enabled and utilized effectively, regardless of environment, while simultaneously informing Quality of Service (QoS) and role-based policy decisions.

The goal of Information Management (IM) is that accurate, relevant, and actionable information be made available to the right consumers. The core IM functions include dissemination, categorization, query, and storage, and should be optimized to the needs of the system's publishers and consumers. The challenge in applying IM principles to semantics is the nature of the data itself becomes fact and relationship based rather than document-based. The constraints of semantic technologies are also distinct from their IM counterparts and require expertise from diverse technological domains: hardware resourcing, storage optimization, semantic reasoning, semantic format, ontology design, and extraction/filter pipeline orchestration. The appropriate use of the IM functions and technologies can enhance the semantic capabilities available to the consumers.

Degraded environments are distinct from enterprise environments because they tend to be deployed as edges to more highly resource provisioned, core environments. In a homogeneous environment, design solutions can be simplified by narrowing the requirement scope to a single set of deployment attributes. Some distinct attributes include bandwidth, processing capacity, policy, federation, transport protocols, node distribution, security, and replication. The core Information Management functions can support widely diverse environments through design adaptation, but require intelligent technology selections and configuration to minimize the impacts upon complexity and performance.

When semantic information is managed in an environment with multiple resource levels, such as a tiered or interleaved enterprise to tactical deployment, the solutions grow in complexity. The growth in complexity is rooted in both the scale and diversity of the environment dimensions.

Environments where applications demand high levels of resources and enhanced performance require a greater scale of hardware infrastructure and / or software complexity. Diversity characterizes the degree of network, hardware, and software heterogeneity. Increasing the diversity of deployment environments corresponds to an increase in the design complexity and performance requirements of suitable technology solutions.

Most technology solutions involve, appropriately, the concentration of tasks within the highly resource provisioned enterprise environments. This concentration results in the reduction of resource demands for processing, storage, or bandwidth within the degraded environment. The causality of these resource demand reductions share a core set of methods: task relocation, replacement, optimization, and elimination. For the most degraded environments, remaining operational can necessitate the migration of nearly all storage, processing, and bandwidth intensive applications to the enterprise level. This can result in the degraded environment possessing only rudimentary tasks and visualization capabilities.

The consequence of our approach is that the applications can be fit to their environment, enabling applications within the degraded environment that were previously too high in resource cost. We identified key milestones to enabling semantic capabilities in degraded environments. The ordered requirements are:

1. Create a scalable middleware implementation with features that flexibly conform to both enterprise and degraded environments.
2. Create an Information Model that supports format, type, and semantic annotations.
3. Relocate the storage, processing, and bandwidth intensive IM features from degraded execution environments to enterprise level services and filters.
4. Create enterprise services to perform semantic extraction, annotation, persistence, and reasoning.
5. Adopt lightweight semantic standards, DOD messaging formats, and extraction technologies for use within degraded environments.
6. Map the expansive, enterprise optimized upper level ontologies to smaller, domain and DOD format specific ontologies for use within degraded environments.
7. Associate extracted semantics and Managed Information Object (MIOs) metadata to semantically defined identities and roles.

As a consequence, degraded environments gain access to low cost semantic information that maintain consistency to mapped DOD formats and is supported by the domain ontologies specific to consumer needs.

MOTIVATION AND RELATED WORK

The goal of our work is to flexibly enable semantic capabilities for both enterprise and degraded environment applications. Each of these environments are suited to different categories of semantic applications. Generally, environments that are highly provisioned with resources serve semantic applications with features for extraction, persistence, reasoning, post-processing of semantic annotations, upper level ontology association, and relationship analytics. Alternatively, degraded environments would ideally co-locate semantic features for simple query results, small scale graphs, semantic metadata and provenance, and visualization.

Our approach involves a combination of Information Extraction (IE), Information Management (IM), Service Oriented Architecture (SOA), semantic reasoning and management, information modeling, Data to Information (D2I), and Quality of Service (QoS) Enabled Dissemination (QED).

Different approaches to IE generally fall into three categories: classifier-based, rule-based, and pattern-based approaches. Classifier-based systems use machine learning techniques to train a classifier that processes a document for extraction words. The classifier determines whether a word should be extracted by considering contextual features associated with both the word itself and of those surrounding it. Examples include Hidden Markov Models (HMM) approaches (Freitag and Mc-Callum, 2000), Relational Markov Networks (Bunescu and Mooney, 2004), and ALICE (Chieu et al., 2003). The rule-based approach to IE uses a set of explicit patterns to find relevant information. Older systems generally relied on manually defined patterns while more recent systems learn them with different degrees of automation. Some examples of these approaches to information extraction are CRYSTAL (Soderland et al., 1995), FASTUS (Hobbs et al., 1997), RAPIER (Califf and Mooney,1999), WHISK (Soderland, 1999), sub tree patterns (Sudo et al., 2003), KnowItAll (Popescu et al., 2004), and predicate-argument rules (Yakushiji et al., 2006).

The result of our work is the production of IM, role, identity, mission, temporal, and geo-spatial semantics which can be used to support IM administration capabilities like authorization and message prioritization. The COTTON model (Tamez et al., 2009) is similar in trying to harness the

resources and capabilities of "helper" devices within opportunistic networks by implementing trust management to improve the security and reliability of the network.

Some pre-existing semantic projects focused on semantic decentralized control and reasoning by implementing an automatically composable rule-based OWL reasoner (Tai et al., 2009) to reduce resource consumption without losing semantic reasoning abilities. Some similar approaches (Jiao et al., 2009) used mobile nodes that relied on a less constrained backbone. OntoMobil (Nedos et al., 2009) attempted to improve discovery of semantically diverse content but use decentralized semantic approach. Other efforts focused on using content fusion based on temporal and spatial ordering of events (Madhukalya, 2012) to reduce redundancy by merging or removing content that was no longer relevant while updating and combining content to improve usefulness of information.

Our research proposes an integration of proven IM solutions paired with novel Information and semantic modeling to enable semantic solutions on previously difficult to manage platforms and environments. Research has pursued the enhancement and maturation of management capabilities within the Publish and Subscribe architectural style, including: scheduling, resource management, policy enhancement, dissemination flexibility, optimum storage schemes, and query paradigms. While these are foundational to our efforts, they do not solve the prevalent issues of semantic relationship-based queries in information management environments with disparate content types and formats.

PHOENIX AIR SERVICES

The Phoenix AIR (Agile Information Representation) semantic middleware was developed as an extension of our Phoenix SOA IM system. A set of pre-existing Phoenix services were adopted as the core IM infrastructure for our data pipeline, and then extended as required.

The Phoenix services are organized into two distinct categories: Edge and Operational. Edge services are fully exposed to edge actors, or may even be located within edge actor devices. Operational services provide IM capabilities and are hosted by remote machines(e.g. cloud deployment) while obscured from consumer interfacing. They can be accessed as required by the SOA for single purpose or orchestrated IM operations.

The base Phoenix Services are either Administrative or Information-based in nature. Administrative services provide functionality that enables advanced information management operations (i.e. authentication and authorization or service brokering). Information services provide the basic functions for managing information (i.e. information type management and information brokering).

The Submission Service (SS) is designed to support the reception of information over Phoenix channels. The SS can host as many or as few input channels as physically possible within hardware and software limitations. The main duty of the SS is to de-serialize and forward information that is received to other IM services such as the Information Brokering Service (IBS), Repository Service (RS), or other SS instances based on the conditions defined by internal policy. The SS may also be configured to perform information validation operations.

Figure 1 - Core Phoenix Services

AS	Authorization Service	**QS**	Query Service
CRS	Client Runtime Service	**RS**	Repository Service
CS	Connection Service	**SBS**	Service Brokering Service
DS	Dissemination Service	**SMS**	Session Management Service
ENS	Event Notification Service	**SS**	Submission Service
FMS	Filter Management Service	**SUS**	Subscription Service
IBS	Information Brokering	**XBS**	Stream Brokering

	Service		Service
IDS	Information Discovery Service	**XDS**	Stream Discovery Service
ITMS	Information Type Management Service	**XRS**	Stream Repository Service

Table 1 - Phoenix Service Acronyms

The Information Brokering Service (IBS) uses a pluggable architecture to support an extensible set of potential expression processor technologies. The actual processing code, specific to the technologies used, is selected at runtime based on the information format, type, and content. One or more SS instances forward the information to the IBS for brokering. The IBS brokers the information, and as a result, tags each information instance with a list of interested consumer channel definitions. These channel definitions are associated with the predicates that matched the information. The IBS then forwards the information instance to a Dissemination Service for delivery.

The Repository Service (RS) has been implemented to support multiple concurrent data stores. The Repository Interface was defined to describe standard, yet extensible, methods for interacting with a data store. This interface is used by the RS as the transparent facade for all data stores, making the RS code 100% reusable amongst data store technologies which are interface compliant. Distinct repository implementations were created using flat file, Mongo, PostgreS, and Berkeley XML DB technologies.

The Dissemination Service (DS) performs simple information distribution operations based on a round-robin scheduling algorithm. The DS is responsible for creating edge information channels to consumers. This service is used by the Information Brokering and Repository Services to deliver information to registered subscribers and query consumers. When an instance of information is pushed to the DS, it retrieves the list of channel definitions from the information instance's resident context and creates the channel(s) if they do not already exist. It then writes the information instance to each output channel. The same instance of information is written to the output channel for each interested consumer, removing the overhead of managing copies.

AIR INFORMATION MODEL

Our publish / subscribe information model is centered upon the concept of a Managed Information Object (MIO) as a data envelope. The traditional MIO is comprised of four elements; payload, metadata, type, and context. Payload consists of the data content. Metadata characterizes the MIO and is used as the basis for matching by the IBS. Type is a reference to the structure of the payload and metadata. Context is used to store any additional descriptive data or attributes about the information instance. Context attributes can be added manually or through extraction within the characterization process.

Phoenix AIR explored and adopted a model of information that flexibly supports both the traditional publish and subscribe constructs, semantics, a multiplicity of metadata and payloads,

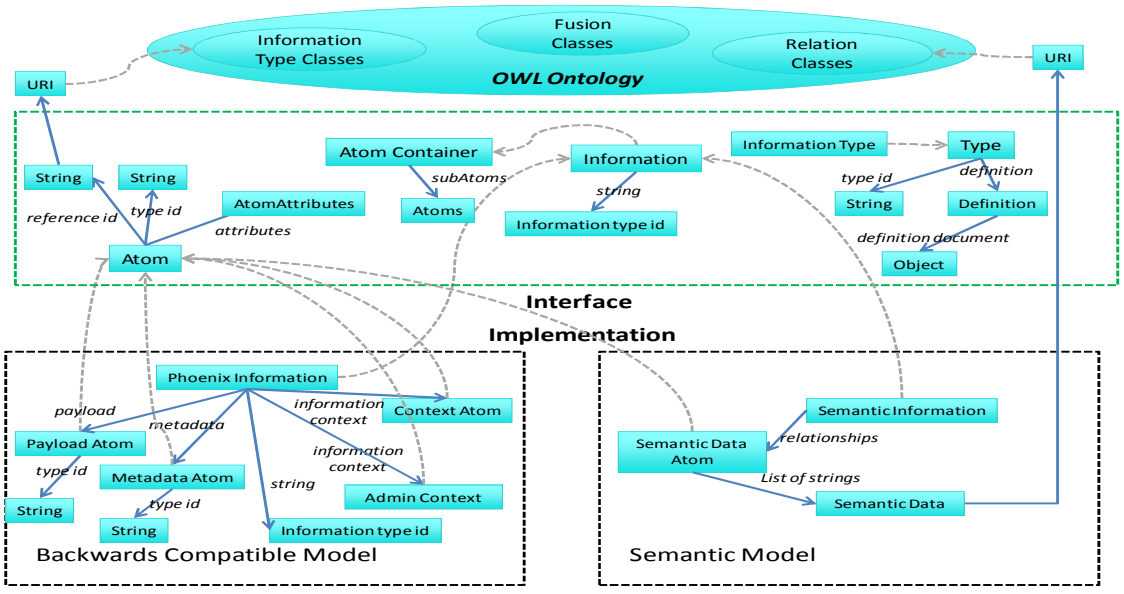

Figure 2 - A simplified visualization of the AIR IM Model

and additional metadata extraction (both semantic and non-semantic) after publication. This information model seeks to answer the lack of flexibility and clarity in the traditional model.

A key enhancement of our semantic enabling model is the creation of an atom-based information container. An atom is a concept popularized by hyper graphs (Iordanov, 2010), wherein an atom database row can have n-ary relationships and a tuple can be defined as the combination of other tuples. An atom is an independent piece of data with its own unique reference id, content, type, and format. An atom type is a reference to a unique, formal structural representation, such as a schema. Content consists of the raw data. A reference ID is a unique identifier of the instance. Some atoms may require only instance uniqueness (Non-ontological, unstructured, or untyped data), while others may require only type uniqueness; type flexibility for semantics allows consumers to adapt the type to application needs.

Atoms are more flexible than either payloads or metadata because it can simultaneously act as a raw or extracted data container, brokering descriptor, or independent data reference, depending upon IBS settings or a possessing service's needs. For our IM purposes, each unique tuple identifier acts as both a reference identifier for independent, non-semantic atom storage and as a semantic Universal Resource Identifier (URI) for Resource Description Framework (RDF) graphs.

The enhanced MIO maintains consistency through convenience methods for payload, metadata, type, and context while adding atom container support via inheritance. As an atom container an MIO can contain an unlimited number of atoms. This breaks the previous 1-to-1 restriction to payloads and metadata. It also lends support, when paired with the new features for mutable atoms and information, for enabling dynamically derived and extracted atoms at runtime.

Another distinction of the AIR model from traditional publish and subscribe models is the type clarity between MIO types and types for payloads and metadata. Independent payload and metadata types create a more flexible model and is truer to reality. Information type has historically been an overridden term that can define multiple concepts:

1. A classification of a general data set (e.g. a higher level schema or ontological class definition). This can be as general as the publisher desires.
2. A unique identifier which acts as a reference to the formal representations (Schemas) of associated metadata and payload of the information.
3. A unique identifier used to define a publish/subscribe 'topic'.
4. A unique identifier which implies the format of the information's metadata or payload.

The new atom type identifiers can be used for the first two given definitions, however, they would improve upon the current implementation by disencumbering types based on the granularity of their mapped schemas. For example, it is technically correct for an information type to be 'CoT', 'USMTF', or 'ship', as they do, in actuality, associate the type with a particular formal structured representation. 'CoT' and 'USMTF' imply a very high level of schema type. It is more correct, however, to associate the concept of type with a more fine-grained schema definition. For example, it is valid for a data's atom type identifier to be *'foaf'* (friend of a friend) rather than *'foaf:document'*. This is comparable to the notion of a type of *'USMTF'*, a course-grained schema, rather than *'MISREP' (Mission Report)*, the more precise, fine-grained schema reference. While coarse-grained, high level schemas are somewhat helpful and valid, types can become more useful when the finer-grained, class-level schemas are also discoverable.

The change of type as a concept is necessary for the pursuit of information model optimization for several reasons. Previously, every type was associated with a schema pairing of metadata and payloads, making the generation of unique pairings overly burdensome to manage. Even if all types were to have singular pairings of payload and metadata, the burden of establishing new types becomes cumbersome where the format of a type may transition from XML (eXtensible Markup Language), but the metadata may now be JavaScript Object Notation (JSON) Format, which is a very distinct concept from type, has also been used interchangeably as the type name. This would previously have required an additional, obtuse definition of an information type for every format change of a schema. Payload or metadata schema names and alterations of a schema format should not result in a cascade of required new types. The new model overcomes these weaknesses through type clarification and independence of types applied to information, atoms, payloads, and metadata.

The four previously defined concepts of an IM system's information type concept are more appropriately classified as follows:

1. Information Type
2. Atom Type
3. Topic
4. Payload/Metadata Format

INFORMATION INDEPENDENCE

Persisting the atom contents of Information objects separately from the Information which possess them is desirable for both performance and the resulting features enabled. Independence of the data content avoids service overhead costs in handling payloads that may never be introspected during brokering or within the service pipeline. It also facilitates the isolation of document-based metadata from data-based metadata.

This separation is critical to semantics because IM principles can apply differently to documents and data. Being independently persisted allow them to inform semantic graphs, while enabling their definition as unique hyper graph tuples. Each atom maintains its own unique identifier because they are intended to be combinational and data-oriented by nature. They should be externally persisted, referenced, and discoverable. As with Information Management, the Semantic Web can be utilized as both document-centric and data-centric dependent upon the technology, applications, and ontology choices. This is evident in contrasting RDF and OWL (web Ontology Language) specifications with their practical application. A unique URI for an RDF subject can be overloaded to refer to either a remote document location or a unique node identifier that references a name, identity, title, or other document-extracted fact.

The lack of a singular focus on either a document-centric or data-centric view of semantics causes great complexity in implementing solutions and results in application specific work-arounds. Phoenix AIR support for the independent retrieval of isolated document payloads or semantic data extractions allows for IM solutions applicable to both paradigms.

SEMANTIC SERVICES

The Phoenix AIR semantic services provide the necessary infrastructure and features to support the creation of consumer semantic applications. The services are designed to assist in providing semantic support regardless of the degradation level of the deployment environment. These services fulfill the semantic roles described in the Introduction to provide resource cost reductions through optimization of persistence, reasoning, and extraction operations. The primary Phoenix AIR services include the Data Management, Characterization, and Semantic Repository Services.

The Data Management Service maintains the format, type, and extraction catalog. Knowledge of available format, type, and semantic extractors are accessible via simple Create, Read, Update, and Delete (CRUD) methods. Each format, type, and semantic extractor possesses a human readable descriptor, although only the extractors contain a list of a list of supported formats and types, and associated ontology dependencies. A Data Management Service user can utilize the extraction descriptor to instantiate the necessary execution code

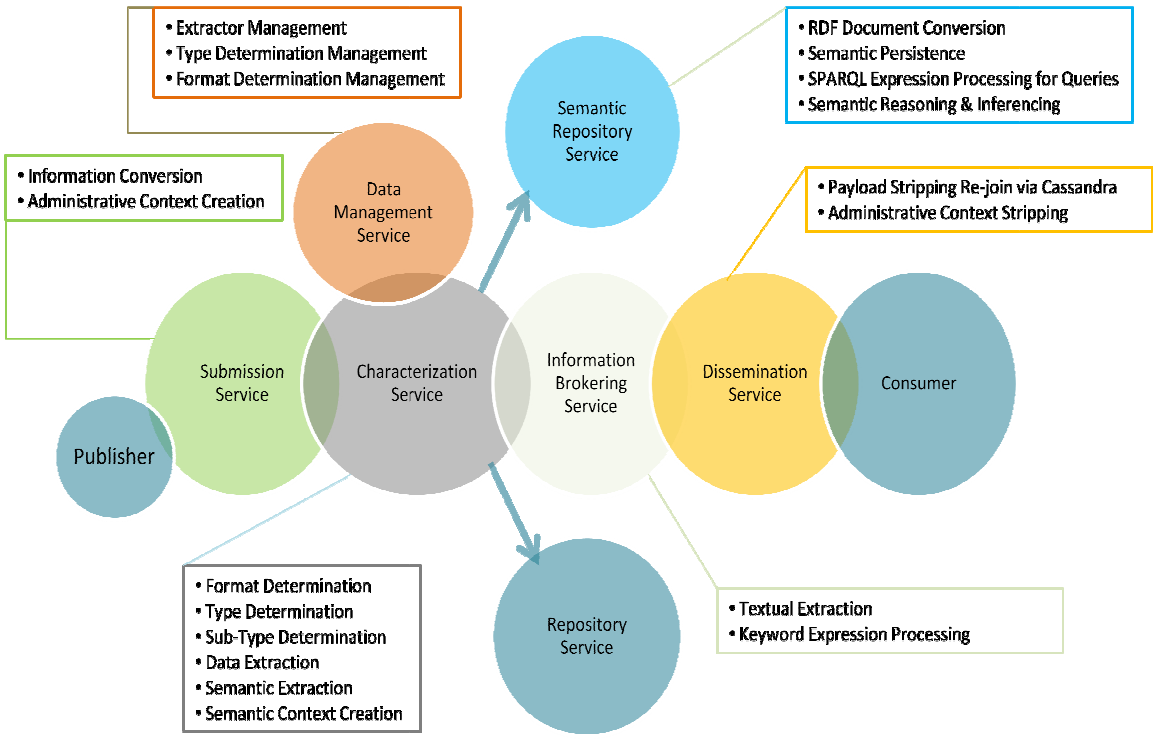

Figure 3 - Phoenix AIR Services

via reflective class-loading.

The Characterization Service utilizes the Data Management Service to perform: autonomous metadata extractions, metadata and payload format recognition, type recognition for unstructured document formats, semantic extractions, and semantic annotations previous to persistence within the Semantic Repository. Upon receipt of an MIO within the Characterization Service, a set of linear processes is set in motion:

1. The MIO data format is determined in-memory via Multipurpose Internet Mail Extensions (MIME) type and magic number recognition libraries.
2. The MIO payload and metadata schema types are determined via a structured introspection via XML and regular expression queries.
3. Based upon the determined formats and types, a catalog of extractors are compiled into an execution list for the MIO.
4. Data extractors are executed upon the payload of the MIO, resulting in a set of atomized documents being added to the context of the MIO.
5. Semantic extractors are executed upon the payload of the MIO, resulting in a set of semantic annotations being generated, compiled, and stored in the MIO context for later persistence within the Semantic Repository.

The Semantic Repository is an Information Management service wrapper of a triple store with methods for semantic queries and reasoning configuration settings. It complies with the Phoenix architecture's interface for services, inheriting the Phoenix default service capabilities for failure recovery, channel management, and service state management.

Many technologies were researched, explored, and prototyped in the pursuit of the AIR semantic IM services. The technologies integrated include:

Function	Technologies
Format Determination	Aperture, Tika, REGEX
Type Determination	Stax, REGEX, XPATH
Semantic Extraction	REGEX, XPATH
Triple store	Jena
Media Player	Google VLC
Graph Visualization	Prefuse
Semantic Inferencing	Pellet, Jena Reasoner
Atom Storage	Cassandra
Keyword Queries	Lucene

Table 2 - Technologies for Semantic Services

SEMANTIC EXTRACTION (SE)

Semantic data extraction occurs within the Characterization Service via the execution of semantic data and type extractors obtained from the Data Management Service. Some are scoped narrowly to a set of types and formats, while others offer general, cross-format, cross-type extractions for time, geo-spatial, or other high level metadata features. Extractors were designed to generate relationships which may be useful in future work enabling information relevancy metrics and influencing role-based and message prioritization policies. A subset of the created extractors consist of:

Temporal Extractor - Extracts any discovered times and dates in the document being processed.

Geospatial Extractor - Extracts anything geographic or spatial in nature from the document, such as latitude longitude points or regions.

Mission Extractor - Extracts the purpose of the mission, the aircrafts involved, the geo-location, the targets, the dates and times of the mission, and point of contact information.

Mission Report Extractor - Extracts the results of the mission, the aircrafts involved, the geo-location, the target results, the finalized date and time interval of the mission, and point of contact details.

ATO (Air Tasking Order) Extractor - Extracts mission ownership, mission tasks, and the dates and times of the tasks.

Identity Extractor - Extracts identity metadata for provenance use, including an MIO's originator, author, and the publisher's current mission role.

In-flight Report Extractor - Extracts the comments, time, and geospatial details of the intermediate reporting of the mission, associating them with the corresponding ATO mission.

Default IM Extractor - Extracts Phoenix IM System specific relationships, such as payloads, metadata, formats, types, and publication date-time.

Imagery Extractor – Extracts semantics temporal, spatial, and imagery provenance such as size, creation time, imagery links, and binary thumbnail.

Engagement Extractor – Extracts information related to targets, tasked units, control points, and initial points.

Map Point Extractor – Extracts relationships regarding the initial and control points for a mission.

Semantic bloat is an issue for SE, with limited options for mitigating the system impact. When an IM system produces a rate of n Information publications per second, but each publication

generates an average of m semantics per Information publication, the scalability limits for the system can quickly be reached. In other words, if an IM system is scalable enough to allow for 10K publications per second, but averages 13 extracted triples per publication, the throughput rate for the IM triple store must be at least 130K triples per second. This form of resource consumption can be mitigated through adoption of highly scalable semantic technologies and intelligent relationship extraction choices.

Establishing the appropriate granularity for a semantic extractor is vitally important to reduce the performance impact of mass generation of semantic data. If granularity is too fine, it can lead to an abundance of obscure and unhelpful domain-based details. Conversely, a granularity which is too coarse can lead to a lot of useless high-level, abstract concepts which are too vague to help user's seeking domain-level query results. The Phoenix AIR SE capabilities enable the optimization for both coarse and fine grained use cases by supporting a pluggable extractor management design.

Once data is extracted and semantically represented, the data can be reused successfully toward three goals: interoperability among missions, decreased disk utilization for storage of duplicate data, and increased depth of knowledge. Since we are reusing data on disk by only storing one instance of a specific piece of data and drawing references to the data, we are able to use those references and their semantic relationships to query, reconstitute messages, and derive new messages. By using the references rather than the actual data, we are able to pass less bytes through the system and retrieve the actual information only when necessary.

ONTOLOGY SELECTION

To express the semantics of everything required, we needed a strong core of ontologies which cover the low level domains and upper level concepts. Our ontology research and selection process spanned medical, social, knowledge organization, and scientific domains, as well as the military and air-force-specific ontologies available. The following ontologies were selected:

U-Core Semantic Layer Taxonomy / Relations - This ontology is a general set of worldview concepts such as vehicles, abstract objects, physical objects, and events. It provides a suitable foundational ontology to create more domain-specific classes and relationships.

Cornerstone Core - This ontology is a core Air Force ontology under development. It contains many concepts which coexist in USMTF and Cursor on Target message formats and their subtypes. It is focused on mission and mission tasking with concepts extended from some of the other ontologies in this list, including the time and U-Core ontology.

Cornerstone Air - This ontology extends the cornerstone core ontological constructs for air mission specific classes and relationships. This enabled us to put the missions in perspective with the Air Force aircraft and UAV (Unmanned Arial Vehicle) resources utilized for the mission scenario.

Time and Temporal - Consists of generic time ontologies for expressing time instances or intervals.

We created some ontologies due to a lack of existing models suitable for our use cases, particularly within the Air Force mission request, reporting, identity, and role domains. The created ontologies include:

Air Force Rank - A simple Air Force Military Rank ontology to enable authentication and authorization for a person's actions based on grade or title.

Air Force Specialty Code (AFSC) - This ontology combined dozens of sources to create a comprehensive ontology for AFSC concepts and properties: career group, career field, career field subdivision, skill level, and specialty code numerical code.

Air Force Tactical Duty Position - An extension of both the Rank and AFSC ontologies, this defines the roles of those in tactical duty positions within the Air Operations Center (AOC) and Air Support Operations Center (ASOC). A producer or consumer's role can be defined and utilized via semantic inferencing to determine related information and desirable message priority.

Information Management - An ontology focused on the Phoenix Information Management system. It expresses common IM terminology for concepts including: publication, subscription, query, information, payload, metadata, and their inter-related properties.

IM Extension - This ontology is a mix of necessary relationships and semantic classes which didn't exist in any of those we discovered. Concepts from this such as 'Target' relationships were created because they did not exist in Cornerstone or U-Core, or did not fit into other ontologies.

Imagery - This ontology represents image concepts and properties including: dimensions, resolution, related side-information, thumbnails, geolocation, and temporal data and other contextual information.

The semantics extraction format is independent of the Phoenix AIR IM, and validated only within the extractors and triple store implementation used. If one extractor outputs semantics in RDF, and another in Turtle, there are no negative system integration consequences or alterations upon the meaning of the data. Serializing semantic annotations via Turtle, however, results in a

much smaller footprint, becoming more optimized and effective for a consumer within a degraded environment.

DATA SCHEMA MAPPING

Mapping from USMTF and CoT formats to ontological concepts provides degraded environments with important data reduction features. A subset of schemas for standard DOD formats, including Cursor on Target (CoT) and USMTF, were mapped to ontological concepts within the Cornerstone and Ucore ontologies. In our scenario testing we decoded an Air Tasking Order (ATO) into its semantics representation: start and end times, missions, targets, aircraft, points of contact, etc. If a consumer were to receive the complete ATO in either its original or semantic format the cost for bandwidth and size of wasted data would be immense. Mapping the XML schema representation to a semantic ontological representation and then performing data reduction processing upon those semantics. The reduction, based upon an extractor or filter is selected by the consumer subscription results in an ideal semantic result for the end consumer.

For instance, if a consumer desires all semantic relationships for Mission 3723, extractors can be put into effect which reduce the published ATO to a simple semantic representation of the single, distinct mission. The resulting semantics are fully compliant with ontological definition of the mission domain concepts, while being traceable to the schema mappings of the original ATO. The unimportant data has been masked and removed from burdening the degraded environment. In our experiments, an ATO with an original size of approximately 294 KB was reduced to 1289 Bytes for the semantic representation of a single mission with 9 ontology dependencies and 16 semantic extractions. When the ontological dependencies were maintained on the consumer rather than the IM services, the size was further reduced to 741 Bytes.

ROLES AND IDENTITIES

The utilization of the tactical duty position ontology provided roles which could be associated with the identity of the publishers of MIOs. Associating identities and roles with IM events and MIOs as provenance metadata, provides the support necessary for making policy and data resource decisions within the IM infrastructure. The importance of provenance for identities and roles is due to their effect upon mission planning and results. This in turn offers better policy for MIO and semantic data queries for degraded consumers.

Our work did not create the policy enforcement points or configuration hooks necessary to empower resource and message management for semantics in degraded environment. It was instead, focused on building the foundation of semantics upon which those features can exist. To make Information Management decisions based upon provenance, roles, identity, and other metadata requires an infrastructure which creates and informs that knowledge. To shape mission and environment resources requires mission knowledge, IM actor knowledge, and enforcements points to be effective.

CONCLUSION

In the course of a single mission, especially one that may have many tasks or correspondence over its duration, there is a staggering amount of messages being transmitted, often to multiple parties. Increase the number of missions and actors operating concurrently, and the amount of raw data that is being transmitted, stored, manipulated, or analyzed increases exponentially. The size of raw data from all of these messages is burdensome to the system with regards to transport bandwidth, storage and memory capacity, and CPU utilization. Semantics can be used to decrease the amount of data through reuse and removal of lower prioritized, redundant, or stale data. However, semantics can also increase the amount of data in the system since the relationships between data, not just the data alone, are also important. Our approach performs end consumer data reduction in multiple ways: Relocation of Processing, Data Extraction, Data Reuse, and Data Referencing.

Due to the infancy of semantic technologies within the degraded or tactical realms, existing ontologies weren't robust enough to express all concepts required to semantically characterize the data and relationships that are commonplace within missions. To mitigate this shortcoming, we extended existing semantic ontologies that were already being used and created new ontologies as needed. Since one of the founding concepts of the Semantic Web is for interoperability, we also made a best-effort to bind concepts within our ontologies to those in other ontologies.

The military domain ontologies, due to their rigidness, are relatively straightforward to extract from. The semantics also require less inferencing support than abstract concepts such as time or knowledge because of their well-defined structure. The greatest weaknesses we found in the ontologies selected were not in their design, but rather in the lack of collaborative and useful concepts.

The adoption of semantic querying capabilities into our system allows the user to query the semantic data to receive the information that is important to them. We are seeking to improve upon raw extractions through the use of semantic

relationships and inference rules to draw connections that may not have been considered by the user but can prove relevancy to their mission. For example, a currently active mission within a given region and time may be unaware of a concurrent mission that is within an acceptable range. Temporally or geospatially overlapping imagery, or other data, could be of interest. The use of inference and reasoning within the enterprise system can lead to the derivation of additional information to determine usefulness. Not only does this improve interoperability and, hopefully, mission success, but all processing is being handled within the enterprise system rather than on resource constrained devices.

Through a combinational solution we were able to reduce the edge processing, bandwidth, storage, and data overhead for degraded environments through the emplacement of appropriate enterprise services, relocation of processing, semantic extractions, data reduction methods, structured semantic schema mapping, and an optimized information model. The results provide a foundation to further work towards relevancy metrics, role-based policy decisions within enterprise environments, and role-based resource and prioritization decisions within degraded environments.

REFERENCES

Iordanov, Borislav. " HyperGraphDB: A Generalized Graph Database". 1st International Workshop on Graph Database, 2010.

D. Freitag, A. McCallum. "Information Extraction with HMM Structures Learned by Stochastic Optimization". 17th National Conference on Artificial Intelligence, August, 2000.

H. Chieu, H. Ng, Y. Lee. "Closing the Gap: Learning-Based Information Extraction Rivaling Knowledge-Engineering Methods". 41st Annual Meeting of the Association for Computational Linguistics, July, 2003.

R. Bunescu, R. Mooney. "Collective Information Extraction with Relational Markov Networks". 42nd Annual Meeting of the Association for Computational Linguistics, July, 2004.

Bryant, J., Combs, V., Hanna, J., Hasseler, G., Krokowski, T., Lipa, B., Reilly, J., Tucker, S., "Phoenix Base Implementation: Final Technical Report", DTIC, 2010.

Arguedas, A., Breedy, M., Carvalho, M., Suri, N., Tortonesi, M., Winkler, R., "Mockets: A Comprehensive Application-level Communications Library", Military Communications Conference, IEEE: Volume 2 (pp. 970 - 976), 2005.

Barroso, L.A., Dean, J., Holze, U., "Web Search for a Planet: The Google Cluster Architecture", Micro, IEEE: Volume 23: Issue 2 (pp. 22 - 28), 2003.

Bryant, J., Combs, V., Hanna, J., Hasseler, G., Hillman, R., Lipa, B., Reilly, J., Vincelette, C., "Phoenix: An Abstract Architecture for Information Management Final Technical Report", DTIC, 2010.

"Command and Control Joint Integrating Concept", Final Version 1.0, DTIC, http://www.dtic.mil/futurejointwarfare/concepts/c2_jic.pdf, 1 Sep 2005.

Schrage, Michael, "The Struggle to Define Agility", CIO Magazine, August 2004.

Shulstad, Raymond A., "Cursor on Target. Inspiring Innovation to Revolutionize Air Force Command and Control", Air & Space Power Journal, winter 2011, http://www.airpower.au.af.mil/apjinternational/apj-c/2012/2012-2/2012_2_08_shulstad-E.pdf.

S. Soderland, D. Fisher, J. Aseltine, W. Lehnert. "CRYSTAL: Inducing a Conceptual Dictionary". 14th International JointConference on Artificial Intelligence, August, 1995.

K. Sudo, S. Sekine, R. Grishman. "An Improved Extraction Patterns Representation Model for Automatic IE Pattern Acquisition". 41st Annual Meeting of the Association for Computational Linguistics, July, 2003.

A. Popescu, A. Yates, O. Etzioni. "Class Extraction from the World Wide Web". 2004 AAAI Workshop: Adaptive Text Extraction and Mining, July, 2004.

A. Yakushiji, Y. Miyao, T. Ohta, Y. Tateisi, J. Tsujii. "Construction of Predicate-argument Structure Patterns for Biomedical Information Extraction". The 2006 Conference on Empirical Methods in Natural Language Processing, July, 2006.

J. Hobbs, D. Appelt, J. Bear, D. Israel, M. Kameyama, M. Stickel, M. Tyson. "FASTUS: A Cascaded Finite-state Transducer for Extracting Information for Natural-Language Text". Finite-State Language Processing. MIT Press, Cambridge, MA. 1997.

M. Califf, R. Mooney. "Relational Learning of Pattern-matching Rules for Information Extraction". 16th National Conference on Artificial Intelligence, July, 1999.

S. Soderland. "Learning Information Extraction Rules for Semi-Structured and Free Text". Machine Learning, February, 1999.

General John Jumper, "Operation Anaconda An Air Power Perspective", February 7 2005, www.af.mil/shared/media/document/AFD-060726-037.pdf.

"Defense Department Special Briefing: Report on the Battle of Takur Ghar", 24 May 2002

Pioch, Nicholas J., Farrell, Robert J., Sexton, William A., Lebling, David, Hunter, Daniel, Barlos, Fotis, "Cornerstone: Foundational Models and Services for Integrated Battle Planning", 17th ICCRTS Operationalizing C2 Agility, March 21 2012.

B. Smith, L. Vizenor, and J. Schoening, "Universal Core Semantic Layer," Proceedings of the conference on Ontology for the Intelligence Community (OIC'09), Oct. 2009.

J. Kopecky, T Vitvar, C Bournez, J Farrell, "SAWSDL: Semantic Annotations for WSDL and XML Schema", http: jacek.cz/publications/2007-11-ieee-ic-sawsdl.pdf, November / December 2007.

D Roman, U Keller, H Lausen, J de Bruijn, R Lara, M Stollberg, A Polleres, C Feier, C Bussler, D Fensel, "Web Service Modeling Ontology", Applied Ontology 1 77–106 77, IOS Press, 2005.

Hongyan Wang; Uddin, M.; Guo-Jun Qi; Huang, T.; Abdelzaher, T.; Guohong Cao; , "PhotoNet: A similarity-aware image delivery service for situation awareness," Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on , vol., no., pp.135-136, 12-14 April 2011

Preuveneers, D.; Berbers, Y.; , "Encoding Semantic Awareness in Resource-Constrained Devices," Intelligent Systems, IEEE , vol.23, no.2, pp.26-33, March-April 2008

Fai Cheong Choo; Seshadri, P.V.; Mun Choon Chan; , "Application-Aware Disruption Tolerant Network," Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on , vol., no., pp.1-6, 17-22 Oct. 2011

Madhukalya, M.; , "Event based content fusion in opportunistic environments," Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on , vol., no., pp.550-551, 19-23 March 2012

Tamez, E.B.; Woungang, I.; Lilien, L.; Denko, M.K.; , "Trust Management in Opportunistic Networks: A Semantic Web Approach," Privacy, Security, Trust and the Management of e-Business, 2009. CONGRESS '09. World Congress on , vol., no., pp.235-238, 25-27 Aug. 2009

Wei Tai; Brennan, R.; Keeney, J.; O'Sullivan, D.; , "An Automatically Composable OWL Reasoner for Resource Constrained Devices," Semantic Computing, 2009. ICSC '09. IEEE International Conference on , vol., no., pp.495-502, 14-16 Sept. 2009

Yazhou Jiao; Zhigang Jin; Yantai Shu; , "Data Dissemination in Delay and Disruption Tolerant Networks Based on Content Classification," Mobile Ad-hoc and Sensor Networks, 2009. MSN '09. 5th International Conference on , vol., no., pp.366-370, 14-16 Dec. 2009

Nedos, A.; Singh, K.; Cunningham, R.; Clarke, S.; , "Probabilistic Discovery of Semantically Diverse Content in MANETs," Mobile Computing, IEEE Transactions on , vol.8, no.4, pp.544-557, April 2009