Realizing Organizational Collaboration through Semantic Mediation
An Approach for Dynamic Data Interoperability within the Intelligence, Surveillance, and Reconnaissance (ISR) Community

Agenda
- Data Mediation Challenges in the ISR COI
- Introduction to Computable Semantics
- Introduction to Semantic Mediation
- Applying Semantic Mediation to the ISR COI
The objective state ISR operational view provides integrated battlespace awareness across multiple data assets regardless of sensor, platform, and organizational boundaries.

- The realization of this vision requires the ability to exchange data in an interoperable fashion in addition to an improved capacity to understand information from a variety of sources.

While the ISR Community has begun to embrace SOA to achieve organization-level information sharing, it has not completely addressed inter-organization interoperability.

- Programs such as the Army’s DCGS-A and the Intelligence Community’s E-Space have embraced Service Oriented Architecture (SOA) concepts
  - Data Services have increased internal visibility and accessibility of data with Web Services and XML technologies
  - Organization-level data interoperability has been achieved through the use of internal data specifications

- Interoperability between DCGS-A and E-Space has not yet been completely achieved due to divergent data specifications
  - Analysts must be able to discover and interpret 3rd party specifications to find external sources of relevant data
  - 3rd party specifications must be mediated to resolve syntactic differences across differing specifications
  - Mediation infrastructure must scale to meet increased demands as the number of available service specifications increases
While Web Services and XML have addressed physical interoperability well, they are still challenged in providing scalable information interoperability solutions

- The core Web Services and XML standards require coded mechanisms to interpret information
  - XML is a platform and application neutral data representation language, but leaves document interpretation up to consumers
  - XSD and WSDL require human intervention to appropriately interpret service capabilities and information requirements
  - XSL/T requires pre-built, hand-coded scripts which only enable syntactic, point-to-point data transformations

- Solutions to these issues have relied on standardized schemas, which do not guarantee cross-organizational interoperability
  - Standardized schemas are difficult to implement
  - Standardized schemas only enforce syntax, not meaning nor usage
  - No single, global schema will meet stakeholder needs across all organizations

Adoption of organization-specific message formats in a purely Web Services and XML world will impact data interoperability across the ISR COI

- XPath and XSL/T provide point-to-point mappings between a single source and a single target

- Point-to-point mappings between COI-specific message formats will not scale
  - N different formats require N^2 – N mappings
  - Modifications to any single schema require changes to 2N – 2 mappings
  - Tightly-coupled, requiring all involved parties to understand how to interpret everyone else’s data

- Tight coupling of XSL/T scripts and mappings violate loose-coupling, a core tenet of Service Oriented Architectures
To embrace true data interoperability, mediation infrastructure must provide the ability to interpret and understand data:

- Information must become the key foundation for organizations and COIs
  - Data are merely physical values
  - Information is a meaningful interpretation of data
- Dynamic information interoperability requires a means to interpret the intention and meaning of data
  - Ability to understand the structure, contents, and business concepts embodied in service contracts and message exchanges
  - Ability to disambiguate the meaning of similarly named terms

An enhanced mediation infrastructure requires an improved ability for software to interpret message formats:

- A loosely-coupled information infrastructure facilitates meaningful interoperability through the use of semantics-based data descriptions
- Semantics-based data descriptions enable a de-emphasis on pre-built, point-to-point mappings
- Mediation infrastructure can transition towards dynamic aggregation and transformation of data by dynamically interpreting data meaning
  - Requires the ability to interpret contents, structure, and meaning of exchanged data
  - Published metadata must describe information contents in an unambiguous, machine-interpretable manner
Achieving Semantic Mediation requires more expressive metadata

- Most forms of metadata focus only on providing syntactic and structural qualities of messages and the services that utilize them.

<table>
<thead>
<tr>
<th>Metadata Type</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntactic</td>
<td>Describes the physical, syntactic markup of individual data elements (formatting, field markers)</td>
<td>Datatype, Field Length, Field Name, Tag Names, Flat File Makers</td>
</tr>
<tr>
<td>Structural</td>
<td>Describes the logical grouping of individual data elements (i.e. entity-attribute groupings)</td>
<td>Logical schema definitions (PersonRecord: PersonName, PersonSSN, PersonDOB)</td>
</tr>
<tr>
<td>Semantic</td>
<td>Describes the codified meaning of data elements, and their relationships, including any rules or constraints on those relationships</td>
<td>Person was-born on PersonDOB, and was-born once and only once</td>
</tr>
</tbody>
</table>

- Semantics is the “meaning of data” – the concepts that data represents within a particular context, and the relationships between those concepts.

Semantics can be formally modeled in an ontology

- An ontology is a graph of the abstract concepts, relationships, and logical assertions that comprise a domain.
  - Usage and meaning of data are explicitly captured in a machine-interpretable format.
  - Machines can automatically discover relevant content sources based on business concepts, not just the static labels currently provided by taxonomies.
  - Ontologies provide a framework for exposing and reusing the interpretation rules coded in currently existing systems.
Ontologies enable software to meaningfully interpret data, lessening human involvement and increasing efficiency

- Ontologies can be used to bridge other models
  - Relationships can be inferred
  - Schema standardization not required

- Ontology constructs can be used to map between ontologies
  - Links are transitive
  - Creates network effect of an enormous scale

Semantic Data Mediation bridges the gap between the data formats and domain knowledge

- XML Schema focuses on describing the proper the syntax and structure of a data format
  - Semantic information is implied, but not explicitly codified
  - OWL provides a rich model to define the semantics of a business domain

- Semantic Data Mediation provides a means to autonomously perform dynamic mediations
  - Semantic mappings provide explicit semantic descriptions of data specifications: Concept Entities, Concept Attributes, and Entity Bridges
  - Two-phased approach allows source XML to be recast in OWL for transformation reasoning and exported into target XML
A Concept Entity is a complex-typed XML element that represents a domain business concept.

Explicitly linked as members of Concept Entities through higher level domain relationships such as hasName(Organization, Name) or hasCity(Location, City).

A Concept Attribute is an XML attribute or element that represents a business domain concept, but has a physical value.
An Entity Bridge represents the higher level domain relationship between two Concept Entities

- Describes how to syntactically and structurally navigate between one the XML element represented by one Concept Entity to another.

Inferencing capabilities allow mediation to occur across data specifications that are not directly mapped

- Transitive nature of ontologies provides implicit bridges between semantic data maps
- Reasoning infrastructure able to infer transformation instruction sets
Semantic Mediation techniques codify implicit knowledge to produce explicit information descriptions

<table>
<thead>
<tr>
<th>Activity</th>
<th>XSLT Mapping/Mediation</th>
<th>Semantic Mapping/Mediation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept Extraction</td>
<td>Implicitly identify the concept entities, concept attributes, and entity bridges in the source and target XSDs</td>
<td>Explicitly document the concept entities, concept attributes, and entity bridges in the source and target XSDs</td>
</tr>
<tr>
<td>Structural Navigation</td>
<td>Implicitly identify the XPath/XQuery to navigate between the concept entities, concept attributes, and entity bridges for both the source and target XSDs</td>
<td>Explicitly document the XPath/XQuery to navigate between the concept entities, concept attributes, and entity bridges for both the source and target XSDs</td>
</tr>
<tr>
<td>Semantic Matching</td>
<td>Manually identify the semantic likeness of concept entities, concept attributes, and entity bridges in the source and target XSDs</td>
<td>Leverage OWL DL reasoning to autonomously determine semantic matching between concept entities, concept attributes, and entity bridges</td>
</tr>
<tr>
<td>Mediation process</td>
<td>For each source to target XSD mediation, manually at design-time compose a stylesheet encompassing the above information</td>
<td>Dynamically at run-time generate a semantic mediation between a source and target XSD</td>
</tr>
</tbody>
</table>

The Semantic Web is a standardized approach towards ontology representation and reasoning that can realize the requirements of Semantic Mediation Infrastructure

- The Semantic Web Activity is a W3C initiative producing standardized mechanisms to specify formal semantics
  - Resource Description Framework (RDF)
  - RDF Schema (RDFS)
  - Web Ontology Language (OWL)
- The Semantic Web stack builds over standard XML and web technologies, easing integration with existing standards
  - Ontologies
  - OWL
  - Taxonomic Categorization
  - RDF, RDFS
  - Schema Description
  - XML Schema
  - Data Aggregation
  - XML Documents
  - Data Values
  - Unicode, URIs, etc.
Compatibility with existing web technologies allows Semantic Web technologies to be integrated into a Service Oriented Architecture implementation:

- ISR service families, in addition to building localized message formats, can build ISR Domain Ontologies encoded in OWL.
- ISR service families describe their XML Schemas using their in OWL.
- Cross-ontology mappings leverage existing mappings to relieve any N² problems.
- ISR organizations register ontologies and OWL-encoded semantic message descriptions.
- Semantic Mediation Service interprets registered ontologies and mappings, performing dynamic mediation and fusion.

With this computable metadata layer, fewer artifacts are required to support information interoperability in the ISR COI:

- Traditional Web Services approach:
  - Organization-proprietary specifications for HUMINT, SIGINT, MASINT data
  - Stylesheet mappings required for each permutation of specification integration and fusion
  - Requires up to 30 mappings
- Semantics-enhanced approach:
  - Create domain ontologies describing ISR domains
  - Single ontology bridge between DCGS-A and E-Space
  - 6 total semantic descriptions, one for each message formats
The semantics-enhanced SOA approach provides a more flexible, scalable mechanism to mediate and consume information.

Traditional Web Services SOA approach:
- XSL/T Mediation Service resolves a point-to-point mapping
- Aggregating from multiple sources requires transforming intermediate results
- Any format change requires 10 mapping modifications

Semantics-enhanced approach:
- Semantic Mediation Service resolves a dynamic mediation routine
- Inferencing over relevant ontologies supports aggregation
- Any message format change requires 1 mapping modification

Semantic Mediation can address a proliferation of ISR-related data specifications in an efficient, loosely-coupled manner

- Provide OWL-backed ontological descriptions for data source schemas and content
- Provide ability to enable dynamic, loosely-coupled any-to-any data transformations and aggregations using a semantics-based mediation techniques
- Complexity grows linearly with the number of different data formats
  - Transitive nature of OWL produces a network effect
- Allows organizations to use data formats tailored for their needs, while seamlessly allowing that same data to be shared with the rest of the community
Questions?

- For additional information or reference materials, please contact:
  - Sri Gopalan, gopalan_sri@bah.com
  - Sandeep Maripuri, maripuri_sandeep@bah.com
  - Brad Medairy, medairy_brad@bah.com

Reference Materials
Ontologies build beyond taxonomy capabilities by providing a codified, machine-interpretable description of a domain

<table>
<thead>
<tr>
<th>Domain Description</th>
<th>Taxonomy</th>
<th>Ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain categorization based purely on keywords</td>
<td>Domain descriptions built through interconnected network of relationships between domain concepts</td>
<td></td>
</tr>
<tr>
<td>Relationships must be assumed: offers no mechanism for describing relationships, sub-type or composition</td>
<td>Relationships are explicit: relationship types between concepts are named, and can be related to other relationships</td>
<td></td>
</tr>
<tr>
<td>The significance of each category name must be understood by the consumer to be meaningful</td>
<td>Offers the relationship types to indicate that differently named terms are equivalent, disjoint, etc.</td>
<td></td>
</tr>
<tr>
<td>Meant as an organizational system for humans to discover and interpret information</td>
<td>Meant as a metadata description framework for machines to interpret information</td>
<td></td>
</tr>
<tr>
<td>Software must be specifically coded against taxonomy category keywords in order to interpret them</td>
<td>Provides rules to interpret relationships and infer new relationships</td>
<td></td>
</tr>
</tbody>
</table>

A semantics-enhanced SOA provides more effective components to realize a traditional Web Services process flow

1. Submit Federated Request
2. Federate Request to E-Space
3. E-Space Response (E-Space Data Specification)
4. Mediation Request (Fuse B and C)
5. Resolve DCGS-A, B, C domain ontologies and mappings
6. Fused Response (DCGS-A Data Specification)
7. Received Federated Response (DCGS-A Data Specification)
Glossary

- **Service-Oriented Architecture (SOA)** - An application architecture approach in which all functions, or services, are defined using a description language and have invocable interfaces that are called to perform business processes.

- **Web Services** - A standardized way of integrating applications using open standards, such as XML, SOAP, WSDL, and UDDI, over an Internet protocol backbone.

- **SOAP** - A lightweight XML based messaging protocol used to encode the information in web service request and response messages before sending them over a network.

- **Web Services Description Language (WSDL)** – An XML formatted language used to describe a web service’s capabilities as collections of communication endpoints capable of exchanging messages.

- **Universal Description, Discovery, and Integration (UDDI)** – A web-based distributed directory that enables businesses to list their services on the internet and discover each other, similar to a traditional phone book’s yellow and white pages.

---

**Glossary**


- **Ontology** – A domain model specifying real-world concepts and their interrelationships. An ontology is typically characterized by non-attributed entities organized not only by subtyping, hierarchical relationships (‘Employee’ is-a ‘Person’), but additionally by semantic relationships describing how one concept is related to another (‘Employee' works-for 'Employer’). Ontologies are commonly used in knowledge representation and artificial intelligence, and are typically used for reasoning, inferencing, and classification computations ([http://en.wikipedia.org/wiki/Ontology_%28computer_science%29](http://en.wikipedia.org/wiki/Ontology_%28computer_science%29)).

- **Semantic Web** – A W3C project creating a standardized mechanism to enable information exchange by giving meaning, in a manner understandable by machines, to the content of documents on the Web. Semantic Web technologies are not limited to Web-centric hypertext media, and can be additionally used to describe the meaning and usage of data and services ([http://w3c.org/2001/sw](http://w3c.org/2001/sw)).
Glossary

› **Resource Definition Framework (RDF)** – An XML-based data model expressing assertions that relate resources (pieces of data) in subject-predicate-object form (RDF Triple). The subject is the ‘thing’ being described, the predicate is the ‘characteristic’ describing the ‘thing’, and the object is the ‘value’ of the ‘characteristic’. This encoding allows software to comprehend sentence-like data assertions ([http://www.w3.org/RDF](http://www.w3.org/RDF)).

› **RDF Schema (RDF/S)** – An RDF-based schema vocabulary language for formally describing groups, or types (known as classes), of RDF resources, and their interrelationships

› **Web Ontology Language (OWL)** – An RDF/S-based ontology language, whose constructs are heavily derived from the DAML+OIL Ontology Language. Adds additional language constructs to provide stronger meaning to RDF/S relationships

› **Reasoning Engine** – A piece of software that attempts to derive answers from a knowledge base. In semantics-based computing, an inference engine typically resolves or discovers interrelationships between ontology classes, allowing conclusions to be drawn about how concepts are related from an underlying ontology.