



**OIC-2007**  
**ONTOLOGY FOR THE**  
**INTELLIGENCE COMMUNITY**  
**TOWARDS EFFECTIVE EXPLOITATION AND**  
**INTEGRATION OF INTELLIGENCE RESOURCES**

**NOVEMBER 28-29, 2007 · COLUMBIA, MD**

**— PROCEEDINGS —**

Editor: Kathleen Stewart Hornsby

*ORGANIZED AND SPONSORED BY:*

NATIONAL CENTER FOR ONTOLOGICAL RESEARCH · BUFFALO  
<http://ncor.us> · [ncor@buffalo.edu](mailto:ncor@buffalo.edu)



135 Park Hall, North Campus · Buffalo, NY 14260-4150 · (716) 645-2444





# OIC-2007

## ONTOLOGY FOR THE INTELLIGENCE COMMUNITY

### TOWARDS EFFECTIVE EXPLOITATION AND INTEGRATION OF INTELLIGENCE RESOURCES

## — TABLE OF CONTENTS —

Dealing with Mistakes in a Referent Tracking System . . . . .	5
<i>Werner Ceusters</i>	
Uses of Ontologies in Open-Source Blog Mining. . . . .	9
<i>Brian Ulicny, Chris Matheus, Mitch Kokar, Ken Baclawski</i>	
Policies for Public Domain Ontologies for the Intelligence Community . . . . .	15
<i>Elisa F. Kendall, Jay Jacobs, Deborah L. McGuinness, Stephen Schwab</i>	
The Use of Ontologies to Support Intelligence Analysis . . . . .	19
<i>Richard Lee</i>	
Creating a Geospatial and Visual Information Ontology for Analysts . . . . .	23
<i>Chumki Basu, Hui Cheng, Christiane Fellbaum</i>	
A Multi-INT Semantic Reasoning Framework for Intelligence Analysis Support . . . . .	27
<i>Terry Janssen, Herbert Basik, Mike Dean, Barry Smith</i>	
Ontologies and Probabilities: Working Together for Effective Multi-INT Fusion. . . . .	33
<i>Eric Little, Kathryn B. Laskey, Terry Janssen</i>	
Ontologies for Rapid Integration of Heterogeneous Data for Command, Control, & Intelligence . . . . .	37
<i>S. Stoutenburg, L. Obrst, D. McCandless, D. Nichols, P. Franklin, M. Prausa, R. Sward</i>	
Ontology-Driven Imagery Analysis. . . . .	43
<i>Troy Self, Dave Kolas, Mike Dean</i>	
Ontological Support for Bayesian Evidence Management. . . . .	47
<i>Michael N. Huhns, Marco G. Valtorta</i>	
Achieving Ontology-Assisted Query of Graph Databases . . . . .	53
<i>David Silberberg, Wayne Bethea, Dennis Patrone, Paul Frank, David Patrone, John Gersh, Elisabeth Immer</i>	
Geospatial Ontology Trade Study . . . . .	55
<i>James Ressler, Mike Dean</i>	
A Pragmatic Foundation for Defining a Rich Semantic Model of Track . . . . .	63
<i>Rick Hayes-Roth, Curtis Blais</i>	
Toward Automated Provability-Based Semantic Interoperability between Ontologies for the Intelligence Community . . . . .	67
<i>Andrew Shilliday, Joshua Taylor, Selmer Bringsjord, Konstantine Arkoudas</i>	
Acknowledgements . . . . .	73



# Dealing with Mistakes in a Referent Tracking System

Werner CEUSTERS

*Center of Excellence in Bioinformatics & Life Sciences, Buffalo, NY*

## 1 Introduction

Referent Tracking (RT) is a paradigm introduced in 2005 intended to provide a means of ensuring unambiguous reference to the particulars in reality that are mentioned in statements of given sorts [1]. Central to this paradigm, which was conceived originally in the context of work on electronic health records, is the use of globally unique singular identifiers – called IUIs (for Unique Instance Identifiers) – that stand proxy for the entities in reality to which they refer. For an identifier (ID) to be a IUI, it must refer to one and only one particular, and this tight connection between the particular and its IUI must be asserted by an author in an RT system (RTS) [2]. One purpose of the RTS is to give agents who wish to make statements about entities in reality a means to retrieve IUIs for particulars to which identifiers have already been assigned, and to create IUIs in other cases. Another purpose is to provide an efficient way to store data about particulars in terms of their relations to other particulars and to the universals which they instantiate [3]. The RT paradigm is associated with developments such as the LSID initiative and the so-called ‘web of things’, but offers a number of advantages brought about by the strict ontological principles under which unique identification is achieved.

## 2 Coping with change

The real world is subject to constant change, and so also is our knowledge thereof. To keep track of these two sets of changes, an RTS requires that any assertion concerning a relationship between entities is associated with an index for the time period during which the relationship obtains, for the time at which the assertion is made, and for the author of the assertion. In [4], we proposed a methodology for realism-based ontology versioning and evolution that makes explicit whether changes in a new version are due to changes in (1) reality ( $\Delta E$ ), (2) the ontology authors’ understanding thereof ( $\Delta B$ ), (3) relevance for inclusion of representations ( $\Delta R_v$ ), or (4) corrections of mistakes ( $R-$  or  $R-$ ). The same methodology can be applied to the RTS because the main difference with a realism-based ontology is that the former is used to store information about particulars, and the latter information about universals. But because of this difference, the RTS will undergo many more changes and this, when used on a large scale, on a constant basis.

## 3 Types of mistakes

### 3.1 *Mistakes in RTS entries*

An entry in the RTS is erroneous if it either violates the principles of referent tracking or fails to mirror the reality to which the RTS is intended to refer.

In either case the entry contains an ID which is believed to be a IUI, but in reality is not. This applies either when the entry does not refer to anything existing (now or in the past), or when it contains a non-singular or non-unique reference. In the case of non-singular reference the same particular is represented in the RTS as two or more numerically distinct entities. In the case of non-unique reference at least two numerically distinct entities are represented as being only one.

Entries in the RTS come in various flavors. One type of entry, called A-tuples (for ‘assignments’), are used to assert the existence of some particular in reality at some time, and to differentiate this particular from other particulars by assigning it a unique singular ID. Hence, A-tuples can be qualified as being in error for one or other of the following reasons:

- A1: the ID does not refer
- A2: the ID refers to two (or more) distinct particulars
- A3: the ID is not the only ID in the RTS that refers to this particular
- A4: the ID does not refer to the intended particular.

A second type of entry, called PtoU-tuples, relates a particular to a universal the reference to which is drawn from some external ontology. Tuples of this type can also be in error for several reasons, including:

- U1: the relationship between the particular referred to by the IUI and the universal in question does not hold during the stated time period,
- U2: the ID for the universal does not refer to the intended universal or it refers to no universal at all.

Where the ID for the particular is subject to an A-type error, the following additional PtoU-errors may occur:

- U3: there is an A1 error in the corresponding A-tuple: the PtoU-tuple is nonsensical
- U4: the ID is subject to a mistake of type A2 and for at least one of the particulars referred to by it, the stated relationship does not hold,
- U5: the ID is subject of a mistake of type A3, and the particular referred to by the ID is not an instance of the universal during the stated time period,
- U6: similar to U5, but involving a type A4 mistake.

Four further types of mistakes are such that reality is mirrored by the PtoU-tuple in question, but what is mirrored is either not what was intended, or is irrelevant:

- U7: the ID is subject of a mistake of type A2, but for *all* particulars referred to by it, the stated relationship holds,
- U8: the ID is subject of a mistake of type A3, but the particular referred to by the ID is an instance of the universal during the stated time period,
- U9: similar to U5, but involving a type A4 mistake,
- U10: there is no A-type of mistake but the stated relationship is irrelevant.

Finally, entries expressed through PtoP-tuples relate particulars to each other and may involve many further sorts of mistakes (P1, P2, ...), depending on whether one or both IDs involve an A-type of mistake, and whether the relationship in question holds.

### 3.2 *Mistakes of omission*

In [4] it is argued that an ontology should contain representational units for all universals that are relevant for the purpose for which the ontology is built. The same principle holds in the context of an RTS: all portions of reality that are relevant for the purpose for which the RTS is maintained should be represented by means of corresponding tuples. If not, the following mistakes occur, in both cases leading to the absence of an A-tuple:

- A-1: the existence of a relevant particular is not acknowledged,
- A-2: the relevance of a particular for the purpose of the RTS is not acknowledged.

Similarly, we recognize two further types of errors involving universals or particulars:

- U-1 / P-1: the existence of a relevant relationship between a particular and some other entity is not acknowledged,
- U-2 / P-2: the relevance of a relevant relationship between a particular and some other entity is not acknowledged.

Whereas mistakes of omission may occur independently of other mistakes, some mistakes of type A and type U will automatically bring in their wake mistakes of other types: Thus for

example, mistakes of type U6 and U9 will be automatically associated with a mistake of type U-1.

#### 4 Dealing with mistakes

To mirror at any given point in time what is believed by the authors of given assertions to be the case in reality at that time, and what was believed to be the case at any earlier point in time, entries in the RTS are never deleted. Rather, the corresponding entities acquire annotations to the effect that they did not mirror reality during the period when they were believed to do so, and possibly also linked to new tuples that function as corrections.

Because of the way an RTS is implemented, no additional templates are required: it suffices to modify the structure of the D-templates as defined in [2] from

$$D_i = \langle IUI_d, T_i, t_d \rangle$$

where  $IUI_d$  is the IUI of the entity registering the tuple  $T_i$  – T symbolizing here any tuple – in the system and  $t_d$  a reference to the time the registration is carried out, to

$$D_i = \langle IUI_d, IUI_{T_i}, t, E, C, S \rangle.$$

This change involves RTS entries becoming assigned IUIs of their own which in the restructured D-template is symbolized by  $IUI_{T_i}$ . The other components of the D-template are:

- $IUI_d$ : the IUI of the entity annotating  $IUI_{T_i}$  by means of the  $D_i$  entry,
- $E$ : either the symbol ‘I’ (for insertion) or any of the error type symbols as categorized in section 3,
- $C$ : a symbol for the applicable reason for change as discussed in section 2,
- $t$ : the time the tuple denoted by  $IUI_{T_i}$  is inserted or ‘retired’, and
- $S$ : a list of IUIs denoting the tuples, if any, that replace the retired one.

We use the Bloodsworth case [5] to demonstrate the principles. In July 1984, 9-year-old Dawn Hamilton was raped and murdered. In August 1984, Kimberly Ruffner was imprisoned for another rape and attempted murder. A composite sketch of the perpetrator in the Hamilton case was shown on the local news. Two anonymous callers advised that Kirk Bloodsworth looked like the composite. Bloodsworth was convicted of the murder. In 1993, new forensic tests discovered semen on Hamilton's underpants. DNA tests proved it was not Bloodsworth's. In September 2003, the DNA sample recovered from Hamilton's underpants was identified as that of Ruffner.

For the sake of conciseness, we describe in Table 1 only verbally what a few relevant IUIs denote, rather than working with a complete ontology. Further relevant tuples not listed in Table 1 are the A-tuples representing the assignment of the IUIs to the corresponding first order particulars, and the D-tuples that go along with them.

Table 2 displays chronologically some of the D- and A-tuples – ignoring their authors – that would result from tracking the particulars. It provides a nice insight into how the RTS changes over time, and how the error correction mechanism goes hand in hand with the representation of changes in reality, our understanding thereof, and changes of relevance. The correction introduced here is the insertion of the D-tuple to which IUI-109 is assigned: this tuple retires PtoP-tuple IUI-9 which contained a Px type of mistake.

IUI-1: Dawn Hamilton	IUI-2: Dawn Hamilton's rape
IUI-3: Composite sketch of Hamilton's rapist	IUI-4: The August 1984 rape
IUI-5: Kimberly Ruffner	IUI-6: Kirk Bloodsworth
IUI-7: the PtoP-tuple representing that IUI-5 committed IUI-4	IUI-8: the PtoP-tuple representing that IUI-6 resembles IUI-3
IUI-9: the PtoP-tuple representing that IUI-6 committed IUI-2	IUI-10: Portion of DNA in Hamilton's underpants
IUI-11: Portion of Bloodsworth's DNA	IUI-12: Portion of Ruffner's DNA
IUI-13: the PtoP-tuple representing that IUI-11 is dissimilar to IUI-10	IUI-14: the PtoP-tuple representing that IUI-5 committed IUI-2

**Table 1: Some relevant particulars and their associated IUIs in the Bloodsworth case.**

<b>Tuple Type</b>	<b>Tuple IUI</b>	<b>Tuple</b>
A	IUI-101	< IUI-1, -, 1975 >
D	IUI-102	< -, IUI-101, July 1984, I, ΔRv, {} >
A	IUI-103	< IUI-2, -, July 1984 >
D	IUI-104	< -, IUI-103, July 1984, I, ΔE, {} >
A	IUI-105	< IUI-3, -, August 1984 >
D	IUI-106	< -, IUI-105, August 1984, I, ΔE, {} >
A	IUI-107	< IUI-6, -, 1961 >
D	IUI-108	< -, IUI-107, 1985, I, ΔB, {} >
D	IUI-109	< -, <b>IUI-9, 1993, Px, ΔB, {}</b> >
D	IUI-110	< -, IUI-14, September 2003, I, ΔB, {} >

**Table 2: some of the D- and A-tuples – ignoring their authors – that would result from tracking the particulars listed in Table 1**

The Bloodsworth case could be represented in many other ways, for instance by assigning a IUI to 'the rapist of Dawn Hamilton' before it is known who that is. In that case, an A3 type of mistake would have to be corrected. The choice of representation is not something that is restricted by the RTS, but rather by the ontologies and the theories upon which they are built.

## **5 References**

- 1 Ceusters W, Smith B. Tracking Referents in Electronic Health Records. In: Engelbrecht R. et al. (eds.) Medical Informatics Europe, IOS Press, Amsterdam, 2005;:71-76.
- 2 Ceusters W, Smith B. Strategies for Referent Tracking in Electronic Health Records. J Biomed Inform. 2006 Jun;39(3):362-78.
- 4 Ceusters W, Smith B. A Realism-Based Approach to the Evolution of Biomedical Ontologies. Proceedings of AMIA 2006, Washington DC, 2006;:121-125.
- 5 Levine S. Ex-Death Row Inmate Hears Hoped-for Words: We Found Killer. Washington Post, September 6, 2003; Page A01.



## Uses of Ontologies in Open-Source Blog Mining

**Brian Ulicny, Chris Matheus, Mitch Kokar, Ken Baclawski**

VIStology, Inc.

Framingham, MA, USA

{bulicny,cmatheus,mkokar,kbaclawski}@vistology.com

### **Abstract**

The blogosphere provides a novel window into public opinion, but its dynamic nature makes it an elusive medium to analyze and interpret in the aggregate, where it is most informative. We are developing new technology employing ontologies to solve this problem by fusing the signals of the blogosphere and zeroing in on issues that are most likely to migrate offline, enabling analysts to anticipate the threats or opportunities they represent.

There are nearly 16 million active blogs on the Internet with more launched every day. Although much of what's discussed in the blogosphere is of little consequence, increasingly, blogs are emerging as powerful organizing mechanisms, giving momentum to ideas that shape public opinion and influence behavior. For example, Malaysian bloggers have recently become quite effective in confronting perceived corruption in their national government despite governmental control of the major media [4]. The blogosphere is thus a great bellwether of changing attitudes and new schools of thought, but only if analysts know which issues to pay attention to and how to identify those issues early in their lifecycle.

Even where there is freedom of the press, blogs provide a more complete picture of public opinion. For example, the New York Times reports that it receives about 1000 letters daily, but publishes only about 15 [1]. By contrast, Google's blog search engine reveals that 3000 or so blog posts on average cite the New York Times every day, many not in English.

### **VIStology's IBlogs Project**

VIStology's IBlogs (International Blogs) project is a three-year effort funded by AFOSR's Distributed Intelligence program to develop a platform for automatically monitoring foreign blogs. This technology provides blog analysts a tool for monitoring, evaluating, and anticipating the impact of blogs by clustering posts by news event and ranking their significance by relevance, timeliness, specificity and credibility, as measured by novel metrics.

Current blog search engines allow users to discover trends in the blogosphere only by determining the most popular names or news articles (e.g. Blogpulse.com) or by overall popularity of the blog itself (e.g. Technorati.com). These metrics favor attention-grabbing stories that may not have lasting significance.

The IBlogs search engine, in contrast, ranks blog posts by their relevance to a query, their timeliness, specificity and credibility. Briefly, these are computed as follows (see [8] for details). In particular, because of the exophoric and quotational nature of blogs, it is important to identify links to news articles that posts cite and analyze them. Blog posts are not standalone documents; therefore, information retrieval metrics must take into account the articles they cite as well as the commentary they add.

**Relevance:** What a blog post is about is determined not only by the text of a post, but also by the text of any news article it references. Terms in news articles and blog posts are not ranked by the familiar  $tf*idf$  metric standard in information retrieval in light of the clumpiness of the corpus and journalistic conventions.

**Timeliness:** The timeliness of a blog post is determined by comparing the timestamp of a blog post with the publication date of a news article that it cites. Timeliness, as distinguished from recency, is about proximity to the relevant event.

**Specificity:** The number of unique individual entities mentioned in a blog post and any news article it cites determines the specificity of a blog post. This is approximated as the number of unique proper nouns and their variants. Attention is also paid to depth in a domain ontology.

**Credibility:** The credibility of a blog author's posts is determined by the presence of various credibility-enhancing features that we have validated as informing human credibility judgments [7]. These include blogging under one's real name, linking to reputable news outlets, attracting non-spam comments, and so on. This analysis must be computed for each author, since blogs can have multiple authors. The number of inlinks alone does not determine blog credibility.

## **Ontologies in IBlogs**

IBlogs uses ontologies and ontological relations in three ways. First, IBlogs uses an explicit domain ontology in OWL for query expansion. Second, IBlogs uses an ontology of the blogosphere to represent and normalize blog data. Third, IBlogs outputs data expressing explicit ontological relations.

Architecturally, the IBlogs systems includes a document extraction module, a metrics computing module, an indexer (Lucene), a crawler (Nutch), an ontology reasoner (BaseVISor) and a consistency checker (ConsVISor). See Figure 1.

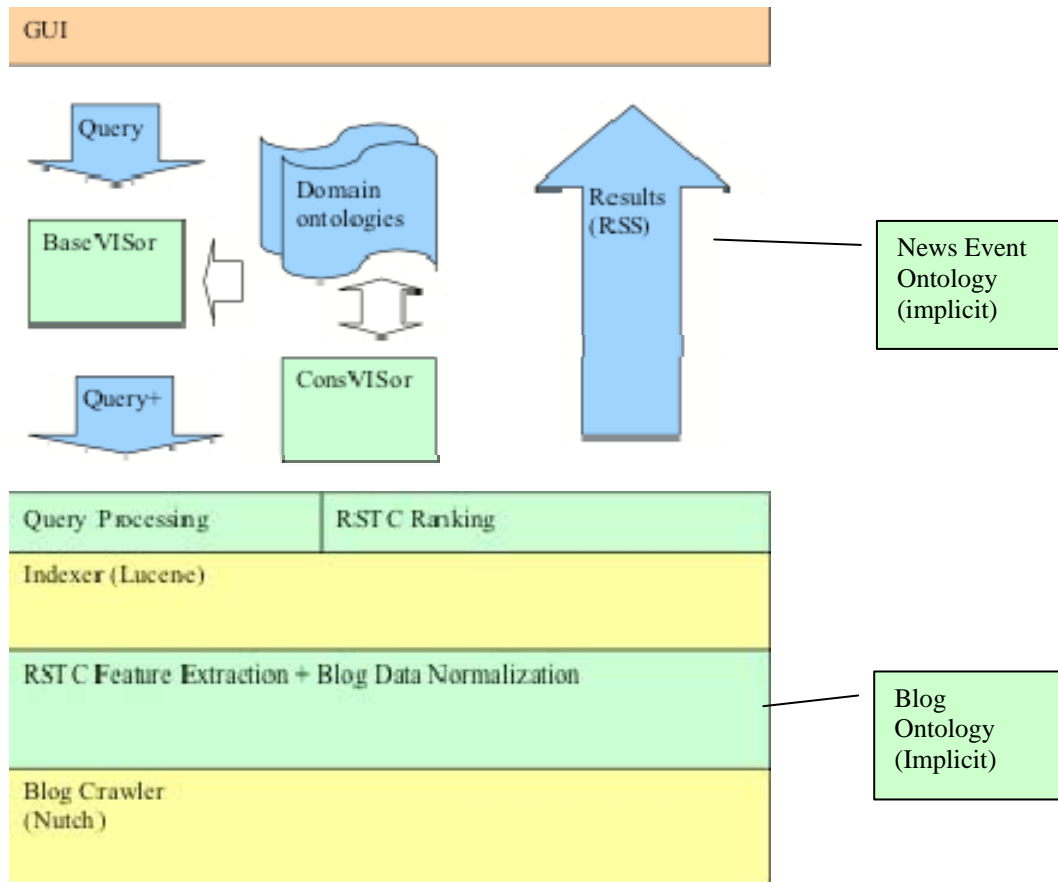


Figure 1: IBlogs Components

VISTology’s BaseVISor inference engine is used to query domain ontologies. At present, we are using a terrorism ontology from Teknowledge. BaseVISor [3] is a forward-chaining inference engine that is based on a Rete network optimized for processing triples. It is able to process RuleML rules containing n-ary predicates, and incorporates the axioms and consistency rules for R-Entailment [6]. BaseVISor allows the system to expand queries based on the domain ontology. For example, the query

[class:TerroristFinancier Damascus]

would be expanded to a query that would return blog posts containing any string that has been included in the domain ontology as a member of the class “TerroristFinancier” and the term “Damascus”.

ConsVISor is a rule-based system for checking consistency of ontologies represented in RDF, OWL, or DAML. We use ConsVISor to help us mediate conflicts and inconsistencies between multiple domain ontologies. ConsVISor can be used to determine whether two entities (with or without the same name) are coreferential [2].

An ontology of the blogosphere is implicit in the system. The Semantically-Interlinked Online Communities (SIOC) ontology [5] provided a useful starting place, but we found it necessary to extend it. While the idea of blogging involves certain essential features, platforms for blogging are not standardized. That is, blogs do not specify what links constitute their 'blog roll', or which links are 'trackbacks' to other blogs, and so on. While feeds for blogs may be specified in several syndication standards (RSS 1.0, RSS 2.0, Atom), these feeds require further analysis because the feed itself is not guaranteed to contain the entire blog post, blog comments, images or profile information relevant to determining blog credibility. All this requires parsing and analyzing HTML blog pages that are designed for human consumption.

Finally, IBlogs outputs information annotated according to an ontology of news events and participants. Our goal is to cluster blog posts by the news events that they are about, where any given news event may have more than one news story that reports it, and each of those stories may be published at one or more URLs. A news event is thus typically two levels removed from a blog post that references it. Our system outputs results in the OpenSearch 1.1 RSS standard (opensearch.org), which we have extended with concepts from the Dublin Core metadata standard (dublincore.org) and with our own namespace elements for news event representations.

NewsML (newsml.org), and the associated EventML standard, represent news industry-originated attempts to standardize representations of news articles and the events they report. These standards can be readily converted to OWL ontologies. We will adapt these emerging standards, currently used by Reuters and Agence France Press (AFP) among others, to standardize the representation of news articles and news events in hope that we will be able to directly use output in these formats produced by news providers in the future.

The IBlogs project demonstrates that ontologies are useful for fusing blog information concerning the elements of the blogosphere, topical subject matter and semantic relations between posts.

## **Acknowledgement**

This material is based upon work supported by the United States Air Force under Contract No. FA9550-06-C-0023. The views and findings expressed here do not necessarily reflect the views of the United States Air Force.

## **References**

1. Feyer, T., *Editors' Note; The Letters Editor and the Reader: Our Compact, Updated*. New York Times, May 23, 2004
2. Kokar, K, C. Matheus, J. Letkowsk, K. Baclawski and Paul Kogut, *Association in Level 2 Fusion*. In Proc of SPIE Conference on Multisensor, Multisource Information Fusion, Orlando, FL., April 2004 (vistology.com/consvisor)
3. Matheus, C., K. Baclawski, M. Kokar. *BaseVISor: A Triples-Based Inference Engine Outfitted to Process RuleML and R-Entailment Rules*. In Proceedings of the 2nd International Conference on Rules and Rule Languages for the Semantic Web, Athens, GA, Nov. 2006. BaseVISor is

freely available (vistology.com/basevisor).

4. Open Source Center, *Analysis: Tension Between Malaysian Bloggers, Authorities Appears To Intensify*, September 13, 2007

5. SIOC Core Ontology Specification <http://rdfs.org/sioc/spec/>

6. ter Horst., H, *Combining RDF and Part of OWL with Rules: Semantics, Decidability, Complexity*. In Proc.of the Fourth International Semantic Web Conference. Y. Gil et al. (Eds.): ISWC 2005, LNCS 3729.

7. Ulicny, B., and K. Baclawski, *New Metrics for Newsblog Credibility*, Proceedings of 1st International Conference on Weblogs and Social Media (ICWSM'07). Boulder, CO.

8. Ulicny, B., K. Baclawski, A. Magnus, *New Metrics for Blog Mining*, Proceedings of SPIE -- Volume 6570 Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security 2007, Belur V. Dasarathy, Editor, 65700I (Apr. 9, 2007), Orlando, FL.



## **Policies for Public Domain Ontologies for the Intelligence Community**

**Elisa F. Kendall<sup>1</sup>, Jay Jacobs<sup>2</sup>, Deborah L. McGuinness<sup>3</sup>, and Stephen Schwab<sup>2</sup>**

<sup>1</sup>Sandpiper Software, Los Altos, CA (ekendall@sandsoft.com)

<sup>2</sup>SPARTA, El Segundo, CA (jay.jacobs@sparta.com, stephen.schwab@sparta.com)

<sup>3</sup>Stanford University, Stanford, CA; affiliation as of date of conference: Rensselaer Polytechnic Institute, Troy, NY (dlm@cs.rpi.edu)

### **Abstract**

Numerous RDF vocabularies and OWL, KIF, and other knowledge representation language ontologies have been contributed to the growing body of ontologies available in the public domain over the last ten years. Many of these were created with government-funded research support in the US and EU. Only a small subset is reusable, and fewer are appropriate for use in applications supporting evolving Intelligence Community requirements. This is partly due to decreasing funding available in the US in particular, but also because of lack of well-specified policies for vocabulary management, metadata, and provenance specification. In this paper we will highlight some of the challenges we have faced in developing and attempting to reuse ontologies in support of DARPA and US Department of Defense initiatives, and provide fodder for discussion of requirements for public domain ontologies.

### **Introduction**

Numerous RDF (Resource Description Framework [1]) vocabularies and OWL (Web Ontology Language [2]), KIF (Knowledge Interchange Format [3]), and other knowledge representation language ontologies have been contributed to the growing body of ontologies available in the public domain over the last ten years. Many of these were created with government-funded research support in the US and EU. Only a small subset is reusable, and fewer are appropriate for use in applications supporting evolving Intelligence Community (IC) requirements. This is partly due to decreasing funding available in the US in particular, but also because of lack of well-specified policies for vocabulary management, metadata, and provenance specification.

Many of the ontologies available from the Protégé library [4], the National Center for Biological Ontology [5], via Semantic Web Central [6], and other collections are domain-specific, focused, for example, on use cases in pharmacogenomics, radiology, or other biomedical or other domain-specific applications. Of those that are more general in nature and potentially relevant for intelligence use, many are incomplete due to funding limitations, reflect varying coverage and granularity, and/or were developed with very specific application requirements in mind. They rarely include the level of metadata and provenance necessary to meet IC requirements [7-8]. Even fewer provide sufficient metadata from a vocabulary management perspective to enable users to understand the ramifications of long-term dependence [9].

Our insights in requirements and methodology for ontology and vocabulary development and management for intelligence use are derived from experience on a number of DARPA, ARDA, other US Department of Defense and NOAA programs as well as commercial projects. They reflect discussions with colleagues in Object Management Group (OMG), World Wide Web Consortium (W3C), and related international standards activities as well as direct conversations

with and surveys of intelligence analysts. And, while individual researchers may have varying opinions on specific aspects of ontology development methodology, choice of language, tooling, and so forth, we have found little to no disagreement on critical issues in vocabulary management or metadata and provenance requirements.

## **Motivation**

A number of the better known, publicly available RDF vocabularies and ontologies, including the OWL language itself and general metadata schemes such as Dublin Core [10] and the Simple Knowledge Organization System (SKOS)[11], were initially created by small teams of developers in collaboration with much larger user communities. It is possible that their utility is responsible for their popularity, but we believe this is also due to the commitment made by the developers to support their users, resulting in continuous improvement over time. In contrast, while the majority of the ontologies developed under the DARPA DAML program are the direct result of significant initial effort on the part of the research community, many of these are showing signs of age and reflect the limited funding available for specific ontology development even over the course of that program. For example, a number of projects, including the time zone ontology components [12] developed for use with DAML Time [13], OWL-S [14], and other domain-specific applications depend on the ontology components for ISO 3166 (codes for the representation of names of countries) available in the DARPA DAML library [15]. This particular ontology provides the set of the alpha-2 codes specified in ISO 3166-1 as of its publication date (2003), but has not been revised since and does not support a number of other data values present in the current standard, such as alpha-3 and numeric codes, references to administrative languages, and so forth. This information was likely not needed when the ontology was initially developed, and some of the detail has been added in a recent revision of the standard. The example highlights issues such as maintaining currency, documenting maintenance policies, describing development requirements, the authority of the publisher with respect to the original standard, and so forth, however, which are clearly important to those who might want to reuse these ontologies in other applications, and particularly for IC applications that clearly must be able to count on currency in this and many other “general” vocabulary subject areas.

## **Vocabulary Management**

The Semantic Web Deployment Working Group has continued work initiated by the Semantic Web Best Practices and Deployment Working Group to publish some basic principles for managing RDF vocabularies and OWL ontologies based on experience with Dublin Core, SKOS, and other ontology development. Some of the most basic issues under discussion include:

- Naming conventions, including use of URIs and publishing ownership and commitments to URI persistence
- Documentation – for example, following the strategies used for Dublin Core, SKOS, and others
- Maintenance policies
- Version management strategies
- Publishing the formal schema (in addition to the documentation)



These represent only the tip of the iceberg, however, in consideration of requirements for utility in IC applications in our view. For certain ontologies, such as those reflecting ISO standards that are published and managed by a formal registration authority, such as the Library of Congress for ISO 639 (language codes) and ISO 3166, we believe that ontology publication should become the responsibility of the registration authority. It is much more likely that members of the IC would trust an ontology published by the registration authority for the standard, or other publicly recognized authority for a particular subject matter (NIST, for example, with regard to units of measure and related standards), than most other potential publishers such as a small company.

Ontology-based applications for operational IC use also require significant metadata reflecting definition provenance, currency, accuracy, completeness, and a development process that is closer to software engineering CMMI-level 3+ compliance than a typical research program would entail.

We believe that from a practical perspective, development of policies for ontology and vocabulary development and management must be established prior to considering development of such public domain resources.

## References

- [1] Dan Brickley and R. V. Guha. RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation, World Wide Web Consortium, Amsterdam, the Netherlands, 10 February 2004. Latest version is available at <http://www.w3.org/TR/rdf-schema/>.
- [2] Mike Dean and Guus Schreiber, eds., Sean Bechhofer, Frank van Harmelen, Jim Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, and Lynn Andrea Stein. OWL Web Ontology Language 1.0 Reference, W3C Recommendation, World Wide Web Consortium, Amsterdam, the Netherlands, 10 February 2004. Latest version is available at <http://www.w3.org/TR/owl-ref/>.
- [3] M. R. Genesereth & R. E. Fikes, Knowledge Interchange Format, Version 3.0 Reference Manual. KSL Report KSL-92-86, Knowledge Systems Laboratory, Stanford University, June 1992.
- [4] Protégé Ontologies Library, 2007. <http://protege.cim3.net/cgi-bin/wiki.pl?ProtegeOntologiesLibrary>.
- [5] National Center for Biomedical Ontology, BioPortal Ontology Library, 2007. <http://www.bioontology.org/>.
- [6] Semantic Web Central, 2007, <http://www.semwebcentral.org/>.
- [7] Pinheiro da Silva, P, McGuinness, D., McCool, R. 2003. Knowledge Provenance Infrastructure. IEEE Data Engineering Bulletin 26(4), pp. 26-32.
- [8] Christopher Welty, J. William Murdock, Paulo Pinheiro da Silva, Deborah L. McGuinness, David Ferrucci, Richard Fikes. Tracking Information Extraction from Intelligence Documents. In *Proceedings of the 2005 International Conference on Intelligence Analysis (IA 2005)*, McLean, VA, USA, 2-6 May, 2005.

- [9] Kendall et al., eds. Basic Principles for Managing an RDF Vocabulary, W3C Semantic Web Deployment Working Group draft, 2007.  
<http://www.w3.org/2006/07/SWD/wiki/VocabMgtDraft>.
- [10] The Dublin Core Metadata Initiative (DCMI) and DCMI Metadata Terms, 2007.  
See <http://dublincore.org/>.
- [11] Simple Knowledge Organisation System (SKOS), 2007.  
See <http://www.w3.org/2004/02/skos/>.
- [12] Feng Pan and Jerry R. Hobbs. A Time Zone Resource in OWL.  
<http://www.isi.edu/~pan/timezonehomepage.html>.
- [13] Hobbs et al., DAML Ontology of Time. <http://www.cs.rochester.edu/~ferguson/daml/>.
- [14] Martin et al., OWL-S 1.2 Pre-Release. <http://www.ai.sri.com/daml/services/owl-s/1.2/>.
- [15] Dean et al., DAML Country Codes. <http://www.daml.org/2001/09/countries/>.

# The Use of Ontologies to Support Intelligence Analysis

*Richard Lee*

Booz Allen Hamilton  
8283 Greensboro Drive  
McLean, VA, 22102, USA  
*lee\_richard@bah.com*

**Keywords:** Information Extraction, Metadata, Ontologies

## 1. Overview

In this paper we describe the Metadata Extraction and Tagging Service (METS) system in use at DIA. We briefly describe the purpose and function of the system. We explain why we chose to use OWL and ontologies rather than simple XML for the representation of the data it produces. We discuss an experiment we conducted on using ontologies for multi-int data fusion. We describe the OWL ontologies we've developed. We conclude with a list of the ontology and data coordination we hope to do in the future.

## 2. Background

A few years ago, we were tasked with evaluating the accuracy and usability of commercial Information Extraction (IE) tools and with determining the benefits of using them to "tag" many years of message traffic.

IE tools process free-text documents and extract from them items of interest. These items can cover a wide range of types of entities (persons, organizations, locations, equipment, dates, etc), and events. It is important to note that the tools do far more than simply identify the presence of such an item in the document – they extract information *about* an item. For a person, this information could include name(s), title, profession, age, hair color, etc. It could also include information about relationships between the person and other entities and events – associates and relations, membership in a group, ownership of

things, instigation of or participation in an event, etc.

We considered the traditional mechanism for XML "tagging" of documents. This consists of placing XML tags around the references to an item in the document, creating XML elements. For example, the Intelligence Community Metadata Standard for Publication (IC-MSP) defines a set of "in-line" tags for this purpose. In the latest version (4.0), it allows for a set of 18 such tags, including a catch-all.

Although the IC-MSP standard does allow for a modest number of attributes, including the xlink set, it was apparent that it – or indeed any representation based on such in-line tags – would be hard-pressed to capture all the useful information produced by IE. Consider the following sentence from a sample document:

"South of Baghdad near the town of Hillah, a suicide bomber blew up his car outside the house of Police chief Maj. Ahmed Suleiman, killing himself and wounding seven, officials said."

While the text indicating specific entities and events can be tagged, all their properties and relationships are another matter:

- owner of car
- owner of house
- occupation, name, title of the intended victim
- agent, location, instrument, victims, etc of the bombing event

- spatial relations amongst the locations

Many of these concerns could be addressed by abandoning the inline-tag representation in favor of a more item-centric representation. This allows for a cleaner and more complete representation of the information, which facilitates discovery and linking of information across data sources. We have therefore gone that route.

However, the lack of a semantic underpinning for XML made us reluctant to use it as the representation for METS data. We wanted to see the data used throughout DoDIIS, across COIs, and we wanted to ensure it could be used to support automated inferencing.

Accordingly, we elected to use an RDF-based semantic representation. Initially, we used DAML (DARPA Agent Markup Language), and then the W3C standard OWL (Web Ontology Language).

### 3. Ontologies

At this time, METS uses a set of three inter-related OWL ontologies which were developed on the program.

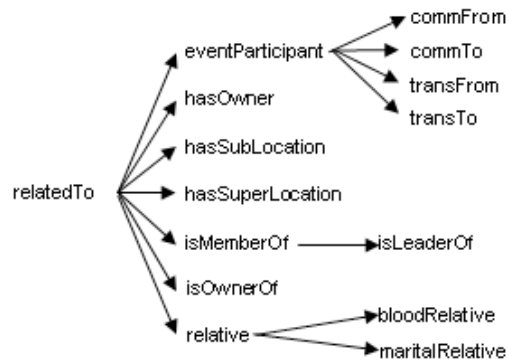
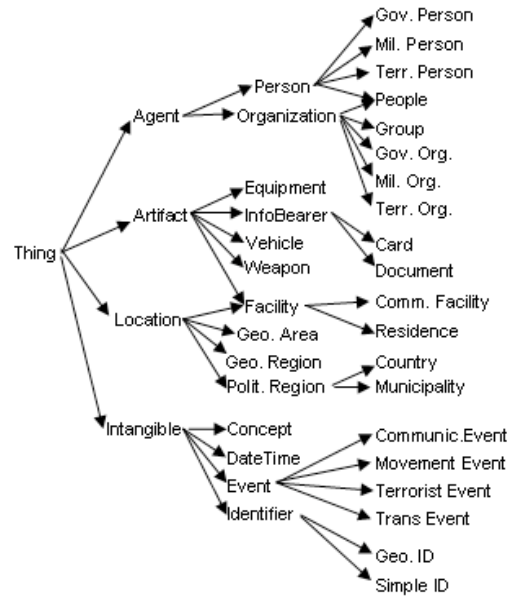
The *core* ontology was designed to arrange a broad set of domain-independent concepts into a class hierarchy. A large set of properties, both for simple text values (name, color, etc) and for relations (memberOf, uses, eventParticipant, etc) was also arranged into a hierarchy. The properties are also identified where appropriate as *transitive*, *inverses* of each other, etc, to further facilitate inferencing.

The *ct* ontology was designed, in like fashion, to cover classes and related properties that were deemed to be specific to the Counter-Terrorism (CT) domain; these were tied into the core hierarchies (via *subClassOf* and *subPropertyOf* declarations).

The *icmsp* ontology was designed to mirror the IC-MSP PublicationMetadata specification,

following its XML structure as closely as possible within the added constraints of RDF. It deviates a bit from that specification to use key items (*Person*, *Organization*, *Date*, etc) out of the core ontology.

Fragments of the class and property (relationship) hierarchies are shown below.



### 4. METS Description

METS is a system for processing text documents. It is fronted by 4 web services:

- *Persistence* service ties to a feed of messages and newswire articles and processes them into a set of data stores
- *On-demand* service accepts arbitrary documents and processes them back to the submitter

- *Query* service retrieves processing results from the data stores matching the query
- *Bulk-transfer* service retrieves all results produced and stored in the specified time interval

METS incorporates a normalization component to convert an input document (text, HTML, XML, word, PDF) into standard XML and OWL forms (see below), and to identify the metadata. It applies a commercial categorization tool and multiple commercial extraction tools, translating the results into the standard forms. It applies heuristics and commercial tools to merge (de-conflict) and clean up the extraction results.

The result of the processing is represented as an OWL/RDF document. All the document metadata (security, date, source, etc information), including the categorization results, is represented in the OWL, using the *icmsp* ontology. All the results of the extraction -- entities, events, and relationships -- are also represented, in conformance with the *core* and *ct* OWL ontologies

Each input document is normalized into XML compliant with the IC-MSP specification. The metadata about the document is represented as called for by the PublicationMetadata portion of the specification. The categories identified by the categorization are included in the IC-MSP metadata as well. The entities and events identified by the extraction are flagged via in-line tags (the set of tags used is actually much larger than the set allowed by the specification, indicating the larger set of entity and event types extracted).

METS is operational at DIA on JWICS, processing live WISE message traffic. Multiple projects are developing interfaces to submit documents and data requests to the METS web services.

## 5. A Multi-INT Experiment

The data processed by METS for storage is message traffic (largely

HUMINT) and newswire articles from WISE. As an experiment, we supplemented the system with a new component which produced OWL from IMINT data, and one which attempted to correlate the data from the two INTs based on location. We enhanced the core ontology with more geographic and geometric concepts to support this; this is of course a prime candidate for carving out and replacing with standard ontologies. The results were encouraging, but suffered from the inability of METS' extractors to disambiguate (and therefore provide coordinates for) location references in many cases.

## 6. Future Ontology and Data Coordination

We will continue to work on improving the coverage and accuracy of the IE in METS.

While the current ontologies were developed in-house, in consultation with CT analysts and their data schemas, we will continue to track and participate in efforts toward standardization such as this conference, work on Catalyst, TWPDES, Universal Core, etc, with the goal of helping devise ontologies that are used and interconnected across the community.

We also hope to be coordinating with other projects to:

- identify coreferential items across the METS-processed documents and other data sources
- discover more knowledge by using the ontology-based inferencing capabilities

## 7. References

Information about METS is at <http://mets.d2lab.net> (internet) and <http://mets.dodiis.ic.gov> (JWICS). The three ontologies are at <http://mets.d2lab.net/onts> (internet) and <http://mets.dodiis.ic.gov/onts> (JWICS).



# Creating a Geospatial and Visual Information Ontology for Analysts

Chumki Basu<sup>1</sup>, Hui Cheng<sup>1</sup>, Christiane Fellbaum<sup>2</sup>

<sup>1</sup>Sarnoff Corporation, 201 Washington Road, Princeton, NJ, 08543

<sup>2</sup>Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08544

## Abstract

*An ontology is a main component of an evolving knowledge base that caters to multiple clients. Consider a scenario where an automated procedure (a computer vision algorithm) used in an analyst tool detects different kinds of “roads” in images, and features in the ontology are used to distinguish a “paved” road from a “dirt road”. In another scenario, the ontology enables reasoning about “locations”, supporting analysts’ geospatial information processing tasks. In this paper, we describe the creation of a multi-use geospatial and visual information ontology, GVIO<sup>1</sup>, building on and integrating with the lexical database, WordNet. To ensure that GVIO can interoperate with other ontologies in useful ways, we inherit as much of the WordNet structure and content as is relevant for the domain of aerial surveillance and link in new content/structure as necessary.*

## 1. Introduction

Geospatial and visual information are essential to intelligence gathering. There is a need to associate meaning with the kinds of entities and relationships useful for information processing tasks (e.g., geospatial query of a region) [1]. In this paper, we describe a Geospatial and Visual Information Ontology (GVIO) we are developing for analyst-specific information processing tasks and computer vision applications. This is a chal-

---

<sup>1</sup> This work was supported by DTO/ODNI under the CASE program. We thank Dan Doney, Emile Morse, and Dennis Moellman for useful discussions. We also thank Glenn Petry at Sarnoff Corporation for his contributions to the reasoning application.

lenging problem as the requirements for these applications can be very different. Whereas an analyst may be interested in locating “facilities near CityX”, the input requirement of a vision algorithm (used in an analyst tool) could be salient “characteristics” of buildings in and around CityX.

## 2. Problem approach

We are interested in understanding how an analyst analyzes the content of aerial video and imagery. There is no single source of knowledge that sufficiently characterizes the information necessary for this type of analysis. Instead, there are a variety of independent resources including WordNet [2], GML<sup>2</sup> (Geography Markup Language), LSCOM [3], Cyc [4], and subject matter experts (SMEs).

We start with the lexical database, WordNet, as our semantic base. Similar to Swartout *et al.* [5], we create an ontology using top-down and bottom-up methods. Our goal is to capture a mix of high, mid-level and domain-specific terms in the ontology, while maintaining the distinction between types and instances defined in WordNet.

### 2.1. Top-down WordNet filtering

We filtered top-level categories in WordNet (Table 1), pruning concepts that need not be further examined (e.g., *Cognition, Food, Feeling* and *Motivation*). We manually classified categories as geospatially/visually relevant, neither, or mixed (relevant and non-relevant). Some categories are mixed and may not be pruned significantly (less than 25%). Other categories (e.g.,

---

<sup>2</sup> <http://www.opengeospatial.org/standards/gml>

*Phenomenon, Causal Agent*) need further inspection to determine the amount to prune (labeled, “undetermined”). Since the distribution of terms across top-level categories is not uniform (e.g., *Event* has a large number of hyponyms), we were left with many unexamined nodes.

## 2.2. Bottom-up data collection

We generated an analyst survey of 400+ terms distilling analyst searches for aerial video/satellite imagery into three lexical categories: *nouns*, *verbs* and *adjectives*. This survey includes an SME concept list for 2/3D computer vision object detection tasks in the urban environment. We list sample terms from each lexical category in Table 2.

We link (map) the terms to *WordNet* synsets. This is a manual step due to polysemy – e.g., we disambiguate the intended sense of “apron”, “a paved surface where aircraft stand while not being used” (ruling out the “protective garment” reading of this word). Rank ordering the terms by respective hyponym tree sizes, we list the top-10 terms in descending order (1). The result is a significant reduction in the total number of relevant or mixed synsets – less than 25% of the total number of synsets in *WordNet*. Combining with top-down filtering, we achieve further pruning (e.g., of terms appearing in the top-level *Person* category).

Person, Location, Tree, Move, (1)  
Leader, Vehicle, Water, Ground,  
Building, Grass

## 2.3. Defining properties

We defined a set of properties for each lexical category. Visual properties of a physical entity are features useful for object detection [6][7]. To maximize utility for object detection algorithms, properties should be quantified, if possible – i.e., assigned default values or ranges of values. For example, we know “telephone pole”, an artifact, has some average “height” based on instances of telephone poles observed. Properties also have

Category	Filtering result
<i>Location</i>	<25%
<i>Event</i>	0%
<i>Act</i>	<25%
<i>Artifact</i>	<25%
<i>Phenomenon</i>	Undetermined
<i>Entity</i>	<25%
<i>Attribute</i>	<25%
<i>Measure</i>	<25%
<i>Cognition</i>	100%
<i>State</i>	Undetermined
<i>Time</i>	0%
<i>Substance</i>	>75%
<i>Relation</i>	>75%
<i>Person</i>	>75%
<i>Communication</i>	>75%
<i>Causal Agent</i>	Undetermined
<i>Possession</i>	Undetermined
<i>Group</i>	<25%
<i>Food</i>	100%
<i>Shape</i>	0%
<i>Natural object</i>	<25%
<i>Feeling</i>	100%
<i>Animal</i>	>75%
<i>Plant</i>	>75%
<i>Motivation</i>	100%

**Table1: Top-down WordNet filtering**

Nouns	Verbs	Adjectives
airfield	carry	armored
barn	chase	barren
hospital	enter/exit	civilian
loading dock	load/unload	dark
telephone pole	meet	rocky

**Table2: Sample terms in analyst survey**

associated subsumption hierarchies – e.g., in Figure 1, “height” is specialized as “sitting height” and “standing height” (useful for pose detection) and default values are assigned for “male” and “female” (derived from anthropometric studies<sup>3</sup>). In Figure 2, hyponyms of “car”

<sup>3</sup><http://ergo.human.cornell.edu/DEA325notes/anthrodesign.html>



inherit “dimension” properties. From *WordNet*, the subsumption hierarchy for “dimension” includes properties, “height”, “width”, and “length”. *CityGML*<sup>4</sup> defines properties for urban settings (e.g., buildings); we link these properties to our ontology, as appropriate.

### 3. Spatial Reasoning from Text

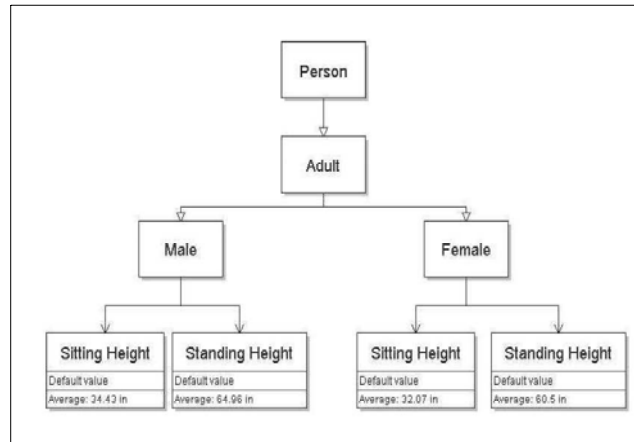
An interesting analyst application that uses ontological relationships is reasoning about spatial entities in text to search imagery/video. Simple keyword-based search is prone to vocabulary mismatches in query terms vs. index terms (from annotations). Often, the annotation space is sparse, resulting in missing data (e.g., location names). Consequently, search queries with location keywords will return no results. A key challenge is the following:

New sites will not be labeled in imagery/video. How do we retrieve imagery/video that contain these locations?

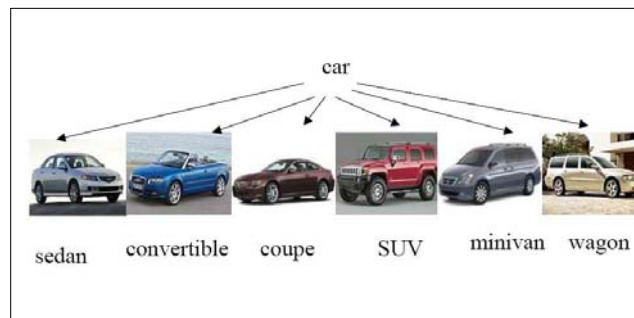
We transform a text-based query (2) into a spatial query expressed in geo-coordinates (latitude, longitude) in multiple steps (Figure 3).

The ABC Training Center is (2) 20 kilometers northeast of CityX.

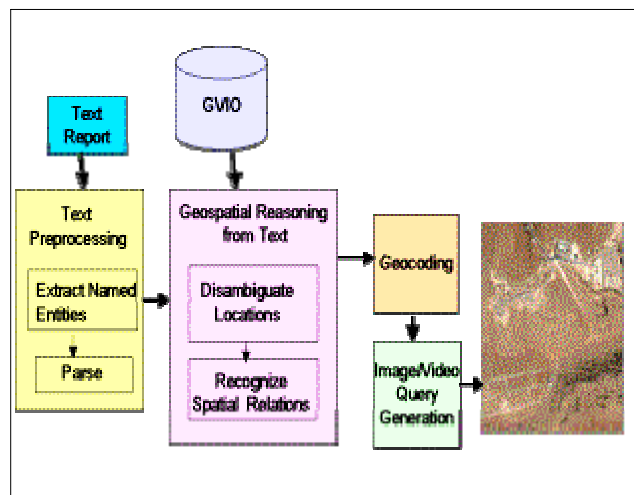
Using a named-entity detector<sup>5</sup>, we find location terms in the original text. We disambiguate a missed detection – “location” incorrectly labeled as “organization” or “person” – using *WordNet*. Using a predefined set of *WordNet* location categories ({WN-Locations}), including “city”, “state”, “country”, “capital”, “lake”, “river”, “building”, etc., and the “instance of” and “hypernym” relationships, we define a function, *inferLocation*:



**Figure 1: Visual property, “height”, specialized as “sitting height” and “standing height” for concept, “person”**



**Figure 2: Types of “car” inherit “dimensions” (“height”, “width”, “length”) as visual properties**



**Figure 3: Flow Diagram**

<sup>4</sup> <http://www.opengeospatial.org/standards/gml>

<sup>5</sup> <http://alias-i.com/lingpipe/>

Direction/Distance	<b>North/South/East/West of Far from, Near</b>
Quantifier + preposition	20 miles <b>From/West of, Very Near/Far from</b>
Simple prepositions	<b>In, On, At, ...</b>
<b>Table 3. Prepositions for analysis</b>	

<b>Configuration1:</b> Location1 is (located, found) Rel Location2
<b>Configuration2:</b> Location1 and Location2 are Rel (near, far, south of each other, ...)
<b>Configuration3:</b> There is Location1 Rel Location2.
<b>Configuration4:</b> Location1 is Rel (south of, far, near, ...) Location2
<b>Configuration5:</b> Ellipsis: Only Location1 is mentioned, Location2 is implied.
<b>Table 4. Spatial configurations in text</b>

```
inferLocation(entity) =
1 if instanceOf(entity) ∈
  {WN-Locations} or
  hypernym(entity) ∈ {WN-Locations},
0 otherwise
```

Prepositions are highly polysemous, which makes disambiguating meaning very challenging [8][9][10]. Table 3 provides a partial list of prepositions/relations to be analyzed. We choose syntactic configurations from the list in Table 4 to disambiguate spatial readings between two locations.

## 4. Conclusions

In this paper, we presented a multi-use geospatial and visual information ontology. We described how object detection algorithms and a geospatial reasoning application benefit from ontology content. We continue to develop this ontology into a general-purpose resource that can be used by analysts.

## References

- [1] M. J. Egenhofer. Towards the Semantic Geospatial Web. *Proceedings of the Tenth ACM International Symposium on Advances in Geographic Information Systems*, McClean, Virginia, November, 2002.
- [2] C. Fellbaum. *WordNet: An Electronic Lexical Database*, MIT Press, 1998.
- [3] M. R. Naphade, A. Hauptmann, S. Chang, and J. R. Smith. A Large Scale Concept Ontology for Multimedia Understanding, Technical Report, 2007.
- [4] D. Lenat and R. V. Guha. *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Addison-Wesley, 1990.
- [5] B. Swartout, R. Patil, K. Knight, and T. Russ. Toward Distributed Use of Large-Scale Ontologies, *Proceeding of the Tenth Knowledge Acquisition for Knowledge-based Systems Workshop*, November 9-14, 1996, Banff, Canada.
- [6] A. Hoogs, J. Rittscher, G. Stein, and J. Schmiederer. Video Content Annotation Using Visual Analysis and a Large Semantic Knowledgebase, in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 2., Madison, WI, 2003.
- [7] R. Srihari and Z. Zhang. Show & Tell: a Semi-automated Image Annotation System. *IEEE Multimedia*, 7(3):61-71, July 2000.
- [8] P. Saint-Dizier (ed.), *Syntax and Semantics of Prepositions*, Springer, 2006.
- [9] W. G. Hayward and M. J. Tarr. Spatial Language and Spatial Representation. *Cognition* 55, 39-84, 1995.
- [10] A. Herskovits. Language and Spatial Cognition: An Interdisciplinary Study of the Preposition in English, *Studies in Natural Language Processing*, Cambridge University Press, 1986.

## **A Multi-INT Semantic Reasoning Framework for Intelligence Analysis Support**

**Terry Janssen (Lockheed Martin), Herbert Basik (Lockheed Martin),  
Mike Dean (BBN), Barry Smith (University at Buffalo)**

### **Introduction**

As is well known, each intelligence agency has developed its own mechanisms for representing data. The resulting stovepipes for transmission pose severe obstacles to the automated integration and management of data, giving rise to problems addressed already in a US Government Executive Order of August 27, 2004, which expressed a mandate to the effect that intelligence agencies must strengthen their mechanisms for the sharing of terrorist information, for example through more widespread and systematic use of XML and similar markup standards.<sup>1</sup>

The volume of available data and the complexity of the National Security environment are increasing so quickly as to overwhelm a finite workforce of analysts. Machines must augment human cognitive capacity in order to achieve the needed level of situational awareness. In what follows, we describe a Lockheed Martin IRAD project to address the problem of integrating the data generated by multiple intelligence agencies. We provide an overview of the project and of the solutions proposed.

Where traditional methods of what is called ‘information fusion’ have been developed primarily for integration of quantitative data, we focus on qualitative data (pertaining for example to *intention or threat*, to *religion* and *family relationships*, or to *relative spatial location*) expressed for example in observation reports.<sup>2</sup> Experience has shown that a combination of semantic technologies is appropriate for capturing such qualitative data. Our goal is to advance the needs of intelligence agents in interpreting very large bodies of such qualitative data by fostering enhanced situational awareness through the application of semantic technology. Little *et al.* describe those aspects of our project which pertain to the use of ontologies to support multi-INT data fusion when enhanced through the consideration of probabilities.<sup>3</sup>

### **Proposed Solution**

The premise of this IRAD project is that a system can be built which allows multi-INT data to be semantically fused and reasoned over by machine.

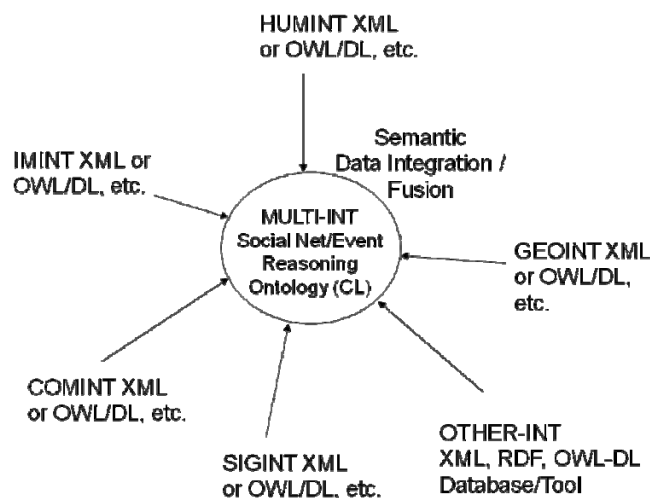
We are building a common framework which provides services to intelligence analysts in a way that does not impose a common vocabulary across the intelligence community or force substantial harmonization of agency-specific approaches to knowledge representation. To this end we exploit the benefits of modularity in building a common upper-level framework to which agency-specific representations can be mapped according to need. The different modules must be interoperable, in order to allow pooling of data from different intelligence agencies. They must also be of high quality in order to gain gradual common acceptance in ways which bring about network benefits of synchronization in the ways in which data are expressed.

The sort of higher-level ontology-based integrating framework we have in mind is being realized already in the context of the Open Biomedical Ontologies (OBO) Foundry initiative.<sup>4,5</sup> Here a plurality of ontology modules is being created by different community groups using both Web Ontology Language (OWL) and OBO-specific ontology formats against a background of common development principles designed to ensure interoperability. The OBO Foundry family of ontologies is being used in large-scale projects for the integration of qualitative biomedical information, including geospatial information,<sup>6</sup> in ways which provide a precedent for the present IRAD project.<sup>7</sup> They provide a set of shared terminological building blocks which foster reliable pooling of more complex representations created in their terms. One crucial component of the Foundry initiative is the availability of reliable ontology converters. These ensure that the large bodies of biomedical data annotated using ontologies (such as the Gene Ontology<sup>8</sup>) created in the OBO format can be transformed into an OWL Description Logic (DL) format. Similar facilities are available to convert OWL-DL ontologies to the CL format within the framework of our present project.

### Semantic Multi-INT Data Integration

Our project hypothesis is that it is possible to create a similar, unified but modular, knowledge space for intelligence-related information integration, comprising both general-purpose open source components (pertaining to geography, religion, transport, etc.) and supplementary special-purpose components provided within various intelligence agencies. The result should provide useful augmentation to human analysts in a way that will help them to achieve the sort of (Level 3) information fusion (situational awareness) which involves integrating characteristics, behavior, political and religious affiliations, locations, etc. of individual entities into higher-level contexts.

Some of the types of IMINT, SIGINT, HUMINT, ELINT, and open source information we will need to integrate are represented in Figure 1. The hypothesis is that even though these different bodies of information are described using different ontologies based on different logical approaches, they can be unified and reasoned over by automated tools given the right sort of computational framework.



**Figure 1: Varieties of multi-INT Information**

## Two levels: the level of classes and the level of instances

Our project rests on a distinction between two levels of entities and of corresponding information that is in some ways analogous to the distinction between T boxes (for terminology) and A boxes (for assertions) used in the DL community. On the higher level are classes or types (and corresponding generic information); on the lower level are instances or particular entities (and corresponding specific information). Classes or types are for example *person*, *settlement*, *plan*, *train*, *aircraft*. Instances are this particular person John or the particular plan that was agreed on by John and four other persons at such and such a time and place. Names of classes or types are used to annotate instance-level data, as for example when a report on image data uses GARCON-F markup to capture the fact that a *business jet* of a certain *type* is *stationary* at a certain *airport*. Ontologies as we conceive them provide the resources to capture generic information in a shareable form which makes associated data themselves shareable and algorithmically tractable.

In the domain of intelligence our information comes from a number of varied sources, many of which will produce either:

- 1) similar information about the same instances,
- 2) similar information about distinct instances (which may accordingly be confused),
- 3) differing information on the same instance (this can produce conflicting information, e.g., concerning the spatial or temporal location of an event),
- 4) differing information about different instances (there may be two separate but related items being tracked in different ways).

Where we are reasoning about how instance-level data fit together to form a common operating picture, there is inevitably uncertainty. Intelligence reports are noisy and information is incomplete. There are active attempts by adversaries at deception. This will mean that all of the mentioned alternatives will generate knowledge problems, for example because we sometimes believe that two instances are identical when they are in fact distinct. In practical terms it means that combining probabilistic reasoning with semantic technology is an important enabling capability for multi-INT fusion. And facilities for probabilistic reasoning will accordingly be an essential component of our project.

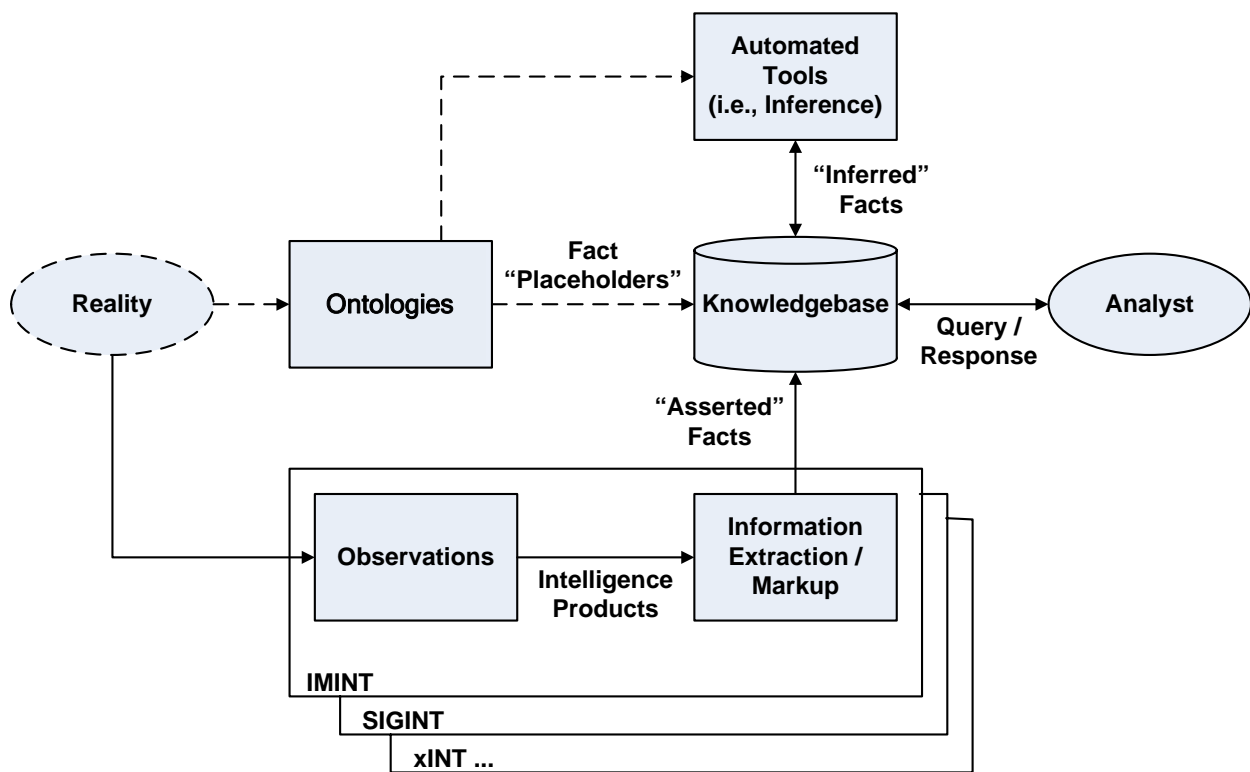
## The Need for High-Expressivity Ontology Languages

Intelligence agencies have developed INT-specific terminologies for describing qualitative data which define terms and relationships in semantically similar but not identical ways. Many of the most advanced of these models have been represented in the OWL-DL format. OWL-DL is a W3C standard with many attractive algorithmic properties. Unfortunately it is a low-expressivity language, which means that it faces considerable difficulties when used to express complex qualitative information especially in areas where time and change are involved.<sup>9</sup> For this reason our project will draw on the resources not only of OWL-DL but also on the more expressive language of Common Logic (CL), a proposed ISO standard.<sup>10</sup> We will draw specifically on the resources of CL with Well-Founded Semantics,<sup>11</sup> which has not only nice computational properties when used in support of reasoning over large bodies of data, but also the sort of high expressivity we need to represent complex real-world situations. CL provides the expressivity we need to describe things that are changing/evolving over time, for example military and paramilitary organizations, family and tribal groups, which gain and lose members, change their

locations, become allied or alienated.<sup>12</sup> Well-Founded Semantics provides fast and efficient query answering capabilities even when addressing large data collections comprehending numbers of entries in the 10s of millions. Like OWL-DL, Common Logic is XML compliant. At the same time CL is marked by a high degree of syntactic flexibility and thus individual CL systems may use a non-XML syntax; these are however in every case mappable to a fully XML-compliant syntax.

Our system will be useful only if it can be executed in responses to real query needs of intelligence analysts in response to real-time changes in real-world environments, and thus it has to be computationally quite nimble. Given that the use of OWL-DL is becoming more widespread it should work well also with OWL-DL resources.

Another key facet of the ontologies of interest to the intelligence community is the ability to express relationships between people, and to construct representations of social networks over populations of individuals. Our project will expand the models of social networks currently in use by adding dynamic spatial and temporal relationships which will be fully integrated within the larger modular framework.



**Figure 2: An Outline of the IRAD Architecture**

## Conclusions

In the foregoing we have described only the basic outlines of the project. In addition we are realizing a number of additional components, including image annotation, data import and results visualization. Our major focus is to construct the engineering required to take this into production (Figure 2), and to bring our pilot testing on artificial data to the level where the approach can be thoroughly tested by information analysts on large bodies of real-world data.

## Acknowledgments

We are grateful to the Lockheed Coporation, and also to the following for crucial contributions to this work: Jim Pridgen, Tony Chen, Jessica Hartnett, Kirsten Gheen, Sarah Taylor, William Andersen, Ryan Kohl, Mitch Kokar, Kathryn Blackmond Laskey, and Eric Little.

## References

---

- <sup>1</sup> <http://usinfo.state.gov/xarchives/display.html?p=washfile-english&y=2004&m=August&x=20040827173509adynned0.6143305>
- <sup>2</sup> Maureen Donnelly, "Relative Places", *Applied Ontology*, 2005. 1 (1).
- <sup>3</sup> Eric Little, Kathryn B. Laskey, Terry Janssen, "Ontologies and Probabilities: Working Together for Effective Multi-INT Fusion" *Ontologies for the Intelligence Community*, Columbia MD, November 2007.
- <sup>4</sup> <http://obofoundry.org/>.
- <sup>5</sup> Smith B, Ashburner M, Rosse C, *et al.* "The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration", *Nature Biotechnology* (in press).
- <sup>6</sup> [http://darwin.nerc-oxford.ac.uk/gc\\_wiki/index.php/EnvO\\_Project](http://darwin.nerc-oxford.ac.uk/gc_wiki/index.php/EnvO_Project).
- <sup>7</sup> Alan Ruttenberg, "Harnessing the Semantic Web to Answer Scientific Questions: A HCLS IG Demo", <http://www.w3.org/2007/Talks/www2007-AnsweringScientificQuestions-Ruttenberg.pdf>
- <sup>8</sup> <http://www.geneontology.org/>.
- <sup>9</sup> Leo Obrst, "Ontologies for Semantically Interoperable Systems", *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, November 2003, 366-369.
- <sup>10</sup> The Common Logic Standard, <http://philebus.tamu.edu/cl/>.
- <sup>11</sup> Weidong Chen, Terrance Swift and David S Warren, "Efficient Top-Down Computation of Queries Under the Well-Founded Semantics", *Journal of Logic Programming*, 1993, 12, 1-199
- <sup>12</sup> K. Hornsby and S. Cole, "Modeling moving geospatial objects from an event-based perspective", *Transactions in GIS*, 11(4): 2007, 225-243.





## **Ontologies and Probabilities: Working Together for Effective Multi-INT Fusion**

**Eric Little (Director, Center for Ontology and interdisciplinary Studies, D'Youville College),  
Kathryn B. Laskey (Center of Excellence in C4I, George Mason University),  
Terry Janssen (Lockheed Martin Corporation)**

Formal ontologies are becoming an essential tool for Intelligence analysis. An ontology provides upper- and domain-level category systems for decomposing and relating objects, object attributes/properties, temporal events, and relations of interest to the intelligence analyst. A great deal of intelligence analysis focuses on understanding and reacting to instance-level report data of varied fidelity on numerous kinds of entities, events and relations. Some of these entities and relations may not be represented explicitly within the ontology's categorical structure, but may be ingested from ancillary systems with which the ontology must interoperate. To an increasing degree, intelligence analysts rely on fusing reports from many different sources. These include reports from different kinds of sensors, processed intelligence from various systems and databases, and human intelligence. The common vocabulary and precisely specified semantics of formal ontologies is a critical enabling factor for interoperability. The promise of multi-INT fusion is that individually noisy and unreliable indicators can be brought together to form a common operating picture (COP) of a given situation [1]. Because the reports being combined may vary greatly in quality, it is essential to account for source quality in combining reports. This requires understanding data quality and applying methodologies for combining information that make use of data quality in a sound and principled manner. Probabilistic reasoning is a well-understood, theoretically sound, and generally applicable method for combining evidence from multiple sources of varying reliability. Computational probabilistic reasoning is a well-established and growing field of research and application (e.g., [2, 3, 4]). Probability has shown its value across a wide range of applications, and many qualitative and heuristic approaches to combining information have been explained as "fast and frugal" approximations to the normative probabilistic solution [5]. Until recently, there has been little research on marrying the fields of formal ontology and probabilistic reasoning. However, this situation is changing (e.g., [6]). This paper will address the question of how formal ontologies can best be combined with probability theory to provide theoretically sound and practically useful semantic technology for multi-INT fusion. We will investigate theoretical concerns associated with the connections between logics associated with formal ontology (e.g., description logic, common logic, first-order logic) and those of probabilistic mathematics. The goal is to provide a high-level discussion of the issues involved with combining ontologies and probabilistic systems as a basis for dialog between these two communities, and to identify a broadly construed research agenda for their mutual development and interaction. The authors of this paper argue the necessity of articulating a clear theoretical foundation as a basis for later development of specific methodologies and languages.

An important question, therefore, is how probabilistic formalisms such as Bayesian Networks can be merged with formal ontologies. Probabilistic theories produce qualified conclusions, graded by numerical measures of plausibility. By contrast, formal ontologies have focused on purely logical reasoning that leads to definite conclusions. Formal ontological categories are

related to one another in definite, law-governed ways, and are understood as possessing a purely binary truth functionality. Formal ontologies are useful for data integration, particularly at the upper-most levels, because they provide for a logical structure for various categories and relations, independent of any particular material knowledge of a given domain [7]. In this sense, a formal ontology provides a means of understanding all types of objects, attributes and relations associated to one another within a given domain by understanding the most basic formal structures they share in common [8].

The question of how formal ontologies can be merged with probabilistic reasoning rests on first defining which items are *in* an ontology per se and which items are *associated with* the ontology (e.g., the reasoning engine, the query language, the results analyzer, etc.). An upper ontology provides asserted facts about the ontic world, meaning the world of ‘general being’ as opposed to any distinctive philosophical or scientific theory of that world – the ontological. So, according to this approach, an upper ontology provides a type of assumed god’s eye view of reality, independent of human observations. By their very nature, human observations presume certain epistemic (i.e., mind- or knowledge-dependent assertions about reality (e.g., as discussed in the lengthy philosophical debates between realist and conceptualist theories of reality) [9-11]. At the upper-most levels, for example, an ontology normally contains non-recursive categorical relations such as: a TerroristAgent is\_a Person, an IED is\_a Explosive, an ObjectShape is\_dependent\_on Substance). The lattice of types and subtypes is a logical structure that is generally taken as given by both logical and probabilistic domain theories. While there may be competing upper ontologies, each with its own type lattice, generally within an ontology, there is no uncertainty associated with the categorical relations.

However, the situation changes when we consider the problem of categorizing instance data. As an example, consider an individual who is declared a Person-Of-Interest, and is being observed to assess whether or not he is engaging in terrorist activities. Information relevant to this problem includes, for example, the network of individuals with whom he associates, his religious affiliation, purchases he has recently made (e.g., materials that could be used to manufacture explosive devices), phone calls to individuals suspected of plotting an attack, etc. The decomposition of Person-Of-Interest into sub-categories of Terrorist and non-Terrorist is a purely logical assumption. However, categorizing an individual as a terrorist or non-terrorist would make use of probabilistic information, such as the base rate of terrorist versus non-terrorist individuals within the relevant population, the likelihood of the pattern of attributes and activities given that the individual is a terrorist versus a non-terrorist, and the credibilities of the reports on which we are basing our inferences about his attributes and activities (e.g., [12]).

Probability is an essential tool for performing this kind of inference in a systematic and principled manner. To perform this kind of reasoning, a system needs the basic categorical knowledge typically encoded in an ontology, and also the likelihood information needed by the probabilistic reasoner. This likelihood information can be obtained from statistical summaries of past instance data, from the judgment of experienced experts, from physical characteristics of sensing systems, or from some combination of the above. This information must be represented in computational form to be processed by probabilistic reasoning algorithms. Increasingly, with the proliferation of distributed fusion systems and web services, it is becoming important to represent this likelihood information not just internally within a given system, but also for

consumption by other systems with which it interoperates. Data quality information must be represented as metadata associated with a web service. When a service returns a result on a situation-specific query, it often must return not just a most likely conclusion, but also information on the uncertainty associated with the conclusion, and also pedigree information to provide the consumer with an audit trail regarding how the conclusion was reached. Interoperating systems require not just shared vocabulary for domain concepts, but also shared vocabulary for communicating statistical regularities pertaining to categories in the ontology, as well as uncertainties associated with instance-level reasoning results, and pedigree information about how conclusions were reached.

An important research issue is how to combine the categorical and relational knowledge typically represented in ontologies with the likelihood knowledge required for multi-INT fusion. Tantalizing to this research agenda is the analysis of how quantitative probabilistic reasoning interacts with qualitatively linked ontological categories. The goal of the current paper is to examine varied approaches to the interactions between ontologies and probabilistic systems, rather than present a clear-cut solution for implementing these kinds of systems (e.g., in multi-sensor fusion applications and the like). Likelihood information must be represented in computational form and combined appropriately with the categorical and relational knowledge contained in ontologies. There have been proposals (e.g., [12]) for augmenting ontology languages to represent probabilistic information. Others (e.g., [13]) have made use of non-probabilistic ontologies to represent structural features of a domain, and have incorporated probabilistic information from outside the ontology to construct a probabilistic model.

To illustrate how probabilistic reasoning can be combined with ontological reasoning, we have developed a simple probabilistic model for multi-INT fusion to identify and head off a potential terrorist attack. The representation language is multi-Entity Bayesian Networks (MEBN) [14]. MEBN Fragments (MFragments) represent small, separable components of probabilistic knowledge about the domain. These MFragments draw on knowledge about ontically existent objects, events and relations, which form *contexts* within which probabilistic reasoning is performed. In the full paper and the presentation, we will describe the case study problem in detail, describe how the likelihood information is used in conjunction with the categorical and relational knowledge, and discuss the question of how to combine probabilistic and ontological technology for problems of this kind.

### **Acknowledgement**

The authors gratefully acknowledge support for this research from Lockheed Martin Corporation.

### **References**

- [1] Hall, D & Llinas, J. (eds.). (2001) *Handbook of Multisensor Data Fusion*, Boca Raton: CRC Press.
- [2] Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA, Morgan Kaufmann.
- [3] Bacchus, F. (1990). *Representing and Reasoning with Probabilistic Knowledge: A Logical Approach to Probabilities*, Boston, MA, MIT Press.

- [4] S. Dzeroski and N. Lavrac, eds. (2001) *Relational Data Mining*. Berlin, Springer-Verlag.
- [5] Martignon, L. and K. B. Laskey (1999). Taming Wilder Demons: Bayesian Benchmarks for Fast and Frugal Heuristics. In *Simple Heuristics that Make us Smart*. The ABC Group, Oxford University Press.
- [6] Costa, P., Laskey, K.B., and Lukasiewicz, T. Uncertainty Representation and Reasoning in the Semantic Web, under review for inclusion in J. Cardoso and M. Lytras (eds), *Semantic Web Engineering in the Knowledge Society*, IRM Press, forthcoming.
- [7] Smith, B & Smith, D.W. (eds.). (1995) *The Cambridge Companion to Husserl* - Cambridge: Cambridge University Press. pp. 27-29.
- [8] Little, E. (2006) "The Need for Metaphysically-based Ontologies in Higher-level Information Fusion Applications," in IFOMIS Report 14 (04/06) Johansson, Klein, Roth-Berghofer (eds.) *WSPI 2006: Contributions to the Third International Workshop on Philosophy and Informatics* (online at: CEUR-WS.org).
- [9] Little, E. & Vizenor, L. (2006) "Principles for the Development of Upper Ontologies in Higher-level Information Fusion Applications," in *Proceedings of the 2006 Formal Ontology for Information Systems Conference*, Baltimore, Maryland, November 9-11.
- [10] Little, E. & Rogova, G. (2006) "An Ontological Analysis of Threat and Vulnerability," in *Proceedings of the FUSION 2006-9th International Conference on Multisource Information Fusion*, July 10-13, Florence, Italy.
- [11] Little, E. & Rogova, G. (2005) "Ontology Meta-Model for Building A Situational Picture of Catastrophic Events," in *Proceedings of the FUSION 2005-8th International Conference on Multisource Information Fusion*, July 25-29, Philadelphia, PA.
- [12] Wright, E. and Laskey, K.B. Credibility Models for Multi-Source Fusion, *Proceedings of the Ninth International Conference on Information Fusion*, July 2006.
- [13] Poole, D. and Smyth, C. Ontologically-Grounded Qualitative Probabilistic Matching, *Proceedings of the Fifth Annual Bayesian Modeling Applications Workshop*, Vancouver, Canada, July 2006.
- [14] Laskey, K.B. (in press) MEBN: A Language for First-Order Bayesian Knowledge Bases. *Artificial Intelligence*.

# Ontologies for Rapid Integration of Heterogeneous Data for Command, Control, & Intelligence

S. Stoutenburg, L. Obrst, D. McCandless, D. Nichols,  
P. Franklin, M. Prausa, R. Sward  
{suzette, lobrst, mccandless, dlnichols, pfranklin  
mprausa, rward}@mitre.org  
The MITRE Corporation

## 1 Introduction

Increasingly Command and Control (C2) systems require the ability to respond to rapidly changing environments and intelligence. C2 systems must be agile, able to integrate new sources of information rapidly for enhanced situational awareness and response to real-time events. Data from varied sources across the world must be integrated and transformed into knowledge that can be leveraged. Machine-to-machine capabilities are also increasingly necessary to accomplish mission goals. To this end, we developed ontologies and rules to address emerging mission needs. We have found that ontologies and rules offer a powerful tool for rapid enterprise integration. With these, we were able to integrate new sources of data within hours, instead of weeks or months as with traditional software development methods. Our work is being showcased at the Joint Expeditionary Force Experiment (JEFX) 2008 for its quick integration of data into usable intelligence-fed C2. This paper describes the use case, the ontologies used to model the use case, and how they support rapid, enterprise integration of C2 and intelligence information, and our prototype Semantic Environment for Enterprise Reasoning (SEER).

## 2 Use Case

Initially our research focused on a military C2 domain with a supply convoy moving through an unsecured area. Figure 1 depicts a convoy moving north along a primary route, approaching the location



Figure 1. Convoy movement using theater, routes, regions of interest (shown as green circle), etc.

where intelligence has reported an enemy sniper is stationed. New information can become available at any time, such as the discovery of a new enemy object in theater, change in

weather, etc., either via immediate convoy recognition or through various intelligence information communicated to the convoy by way of intelligence summaries (INTSUM) and visual and ground moving target indications (VMTI and GMTI).

Both sources of military intelligence, INTSUM provides a summary of the most current enemy situation covering a period of time designated by the commander whereas GMTI/VMTI provides real time information on ground movers. Both are the result of human reported and sensor based intelligence. Through the ontologies and associated rules, the system provides alerts and recommendations to the convoy commander. The alerts and recommendations enhance situational awareness by fusing events; that is, multiple events from different intelligence sources are combined to form battlefield conditions, which trigger alerts and recommendations. In Figure 1, a convoy has moved so that now its region of interest (the circle surrounding the convoy) has encompassed an enemy unit. In this situation, the system might generate an alert based on an intelligence report of enemy sniper in the vicinity and recommend that the convoy take an alternate route [1].

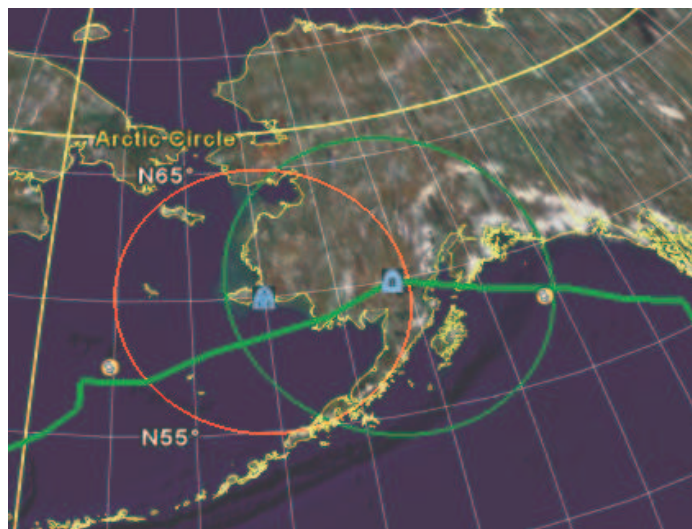


Figure 2. A pilot enters an area of degraded satellite communication. The ROI in red shows the projection of the satellite coverage area onto the Earth.



Figure 3. Google Earth view showing constellation of satellites in real time (satellite positions obtained from WWW).

After showing how ground position and intelligence data could be integrated using ontologies, we extended our prototype by adding event types, including space events, live satellite positions and ship movement, as reported by additional intelligence sources. We added these events in just hours. As an example, Figure 2 shows a pilot entering into an area in which satellite communication is degraded. Figure 3 shows a constellation of satellite positions.

### 3 Ontology Design

To model the objects and events described in Section 2, we constructed five ontologies:

- TheaterObject – battlefield objects and reports about them.
- RegionOfInterest – battlefield regions of interest.
- Convoy – the convoy, its mission, components, etc.
- Convoy Routes – routes the convoy might take.
- ConditionsAndAlerts – how the knowledge base aggregates events, resulting in conditions and alerts that affect the convoy.

Figure 4 shows the high level relationships between each original ontology and its major concepts (in blue and red; subsequent modifications are in yellow). TheaterObjects are MilitaryUnit, Sniper, RoadObstacle, and Facility. TheaterObjects have a location, and may have a speed, heading, and combatIntent (hostile, friendly, etc.).

To distinguish the entity in theater from reports about it, we specified the class of ObservationArtifacts, intelligence reports about objects in theater. ObservationArtifacts have properties such as timeOfObservation, locationOfObservation, speedObservation, etc. The distinction between theater object and observation is important, allowing inference over multiple reports about the same object in theater.

The RegionOfInterest (ROI) ontology models the geospatial areas of special interest surrounding TheaterObjects. An ROI is centered on the position of its focal object, and has shape, dimensions and area -- the dimension and area dependent on the type of threat or interest. ROIs are used to define a “safety zone” around a convoy which must not be violated by hostile or suspicious objects. An ROI also models the area around a reported hostile track that defines the potential strike area of the threat.

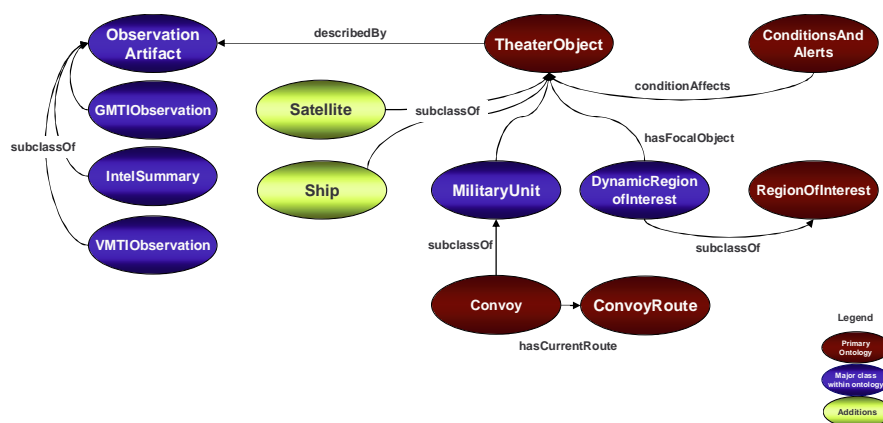


Figure 4. Original ontologies, with modifications (in yellow).

The Convoy ontology, using [2], models the organized blue (US & allied) forces moving on the ground and includes the Convoy's mission, components and personnel.

The ConvoyRoute ontology represents the paths of a convoy, including critical points (CPs) for primary and alternate routes. Recommended routes can change based on application of rules.

The ConditionsAndAlerts ontology models situations on the battlefield based on aggregations of events and actions of theater objects. Conditions based on events result in alerts and recommendations to blue forces.

With SEER we are able to provide new capabilities very quickly. For example, by adding satellite positions and maritime events (displayed in yellow in the figure) to the TheaterObject ontology, instances of those classes are automatically retrieved. We are thus able to integrate new sources of data in hours.

#### 4 SEER Prototype Design

We integrated the ontologies and rules that model C2 scenarios and battlefield intelligence into a loosely coupled service-oriented architecture that uses XML-based messages. The high-level design of the application is shown in Figure 5. The components of the system include the following.

- Enterprise Service Bus (ESB)
- Google Earth<sup>1</sup> Client
- SWORIER (Semantic Web Ontologies and Rules for Interoperability with Efficient Reasoning) [3]:
  - Reasoner, implemented in AMZI! Prolog Logic Server<sup>2</sup>
  - Knowledge Base (KB), composed of ontologies in OWL with instances, rules in SWRL
- Situational Awareness Service (SAS)
- Event Mediation Services (EMS)
- Adaptors
- Message Simulator (MS)

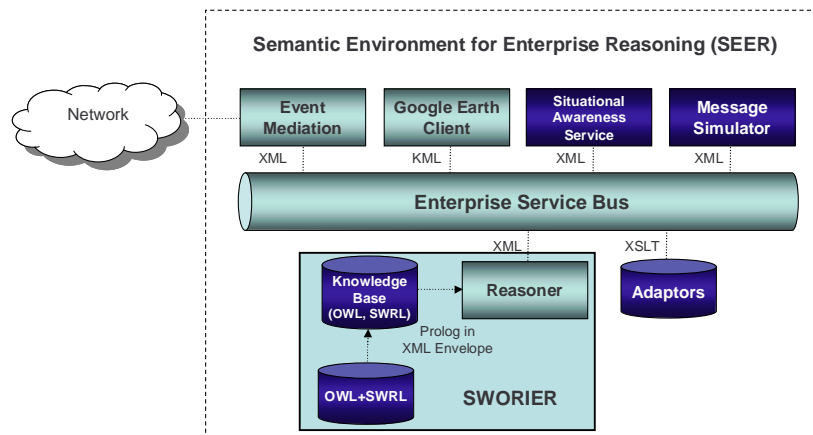


Figure 5. SEER architecture with SWORIER.

We use Mule<sup>3</sup> as the ESB abstraction layer over disparate messaging technologies, allowing interaction between components with minimal code development. Mule supports

<sup>1</sup> <http://earth.google.com/>

<sup>2</sup> <http://www.amzi.com/>

<sup>3</sup> <http://mule.codehaus.org/>



transport and transformation of publisher/subscriber pairs, applying the XSLTs of the Adaptors when appropriate. Mule also detects events, including trigger events that cause the swapping of knowledge bases, enabling us to integrate sources for satellite information and other events.

We use Google Earth since it offers seamless integration of multiple data sources via its Keyhole Markup Language (KML), and also provides excellent maps and zoom capabilities.

AMZI! is the platform on which we host the integrated ontologies and rule base to perform efficient runtime reasoning.

The KB consists of integrated ontologies, rules and instances. OWL ontologies and SWRL rules were translated to Prolog, then optimized [3]. Together with the reasoner, these constitute SWORIER.

SAS detects events (message exchanges over the ESB), consults the knowledge base, and delivers appropriate alerts and recommendations to the convoy commander via Google Earth clients. Events can be object movement, changes in weather, changes in alert conditions, etc. These events constitute reception of simulated INTSUM, GMTI, VMTI, and other intelligence reports. The service can dynamically query the KB.

EMS handles different types of service communication including SOAP synchronous request/response, SOAP pub/sub, polling and REST. SEER uses EMS to interact with outside message sources.

The Adaptors are a set of XSLTs that are invoked by the ESB to translate messages to the appropriate format as they move between components. Events are in an XML format that contains the AMZI! command format, and are asserted to the KBs and translated to KML for display on Google Earth. The active KB generates alarms and recommendations (when queried by the MS) and these messages are translated to KML for display.

The MS sends messages over the ESB to simulate events on the battlefield.

The SEER application works as follows. First, messages are received on the ESB, either from network sources or by the MS. The ESB applies the appropriate XSLTs of the Adaptor and commits the new information to the KB and sends KML to Google Earth.

## 5 Conclusion

Ontologies can be applied for rapid enterprise integration, allowing delivery of new capabilities for example in C2-Intelligence applications in hours, as long as a clear distinction is made between intelligence information reception and actual theater object representation and behavior.

## References

1. Stoutenburg, S.; Obrst, L.; Nichols D.; Samuel, K.; and Franklin, P. 2006. Applying Semantic Rules to Achieve Dynamic Service Oriented Architectures. *Second International Conference on Rules and Rule Markup Languages for the Semantic Web, at ISWC 2006*, Athens, GA, November 10-11, 2006, LNCS 4294, 2006, p. 581-590. Heidelberg: Springer.
2. Convoy Operations Battle Book. 2005. Tactics, Techniques, and Procedures for Training, Planning and Executing Convoy Operations. Tactical Training and Exercise Control Group. Marine Aviation Weapons and Tactics Squadron One.
3. Samuel, Ken; Leo Obrst; Suzette Stoutenburg; Karen Fox; Adrian Johnson; Ken Laskey; Deborah Nichols; and Jason Peterson. 2007. Applying Prolog to Semantic Web Ontologies & Rules: Moving Toward Description Logic Programs. *The Journal of the Theory and Practice of Logic Programming*, Massimo Marchiori, ed., forthcoming.



# Ontology-Driven Imagery Analysis

Troy Self, Dave Kolas, Mike Dean

BBN Technologies

1300 N. 17<sup>th</sup> Street, STE 400

Arlington, VA 22209

[tself@bbn.com](mailto:tself@bbn.com), [dkolas@bbn.com](mailto:dkolas@bbn.com), [mdean@bbn.com](mailto:mdean@bbn.com)

## Abstract

*This paper presents a new paradigm for imagery analysis where imagery is annotated using terms defined in ontologies, enabling more powerful querying and exploitation of the analysis results. The ontology terms represent the concepts and relationships necessary to effectively describe the objects and activities within a domain of interest. A platform for viewing and editing imagery annotations is described along with a specialized semantic knowledge base capable of efficiently querying the information using semantic, spatial, and temporal qualifiers. The ontologies used for representing the annotations and domain of interest are also described.*

## Introduction

Imagery analysis is the process of examining overhead imagery, identifying the objects and activities present in the image, and correlating this data with information not available in the image to derive new knowledge. Current practices for capturing imagery analysis results as narrative text or in relational databases become a burden when an analysts needs to search past reports, correlate facts across multiple reports, and search for specific examples of general scenarios.

The purpose of this paper is to describe an imagery analysis environment where observations are recorded as structured annotations using descriptive semantic concepts defined in an ontology, enabling more powerful search and exploitation of the annotations than can be achieved using traditional methods. The first section describes the user environment for the ontology-driven imagery analysis application. The second section describes the specialized knowledge base developed to enable efficient storage and retrieval of the annotations. The third section describes the ontologies developed to achieve the goals of the application.

## Imagery Analysis Environment

The imagery analysis application is implemented as a plugin for the ESRI ArcMap GIS [1], a popular image analysis tool. The plugin includes a custom layer for viewing geo-registered imagery and marking annotations as well as custom user interface controls for creating new and searching existing annotations. The user interface controls' content is generated dynamically based on the ontology terms and relationships in the knowledge base. This allows users to immediately leverage modifications and enhancements to the

domain ontology without having to wait for deployment of new version of the application.

The user creates new annotations by using the custom controls to describe the observation using the semantic terms defined in the domain ontology. The application automatically captures the timestamp and geospatial details of the annotation. By capturing the temporal and spatial extent of the observation, annotations can be linked and searched using time, space, and description regardless of whether they originated from one or more images.

When viewing imagery, the user can use the custom query controls to filter the visible annotations based on spatial, temporal, or semantic qualifiers. For example, when viewing an airport, the user can choose to only show observations of support vehicles within the hangar area within the past 7 days. This search relies on the ability of the knowledge base to understand what qualifies as a support vehicle and to efficiently eliminate observations that occur outside the specified spatial and temporal extent.

The user environment also includes an advanced query interface that allows the user to write custom queries that cannot be defined using the UI controls. As an example, this interface allows the user to query for all cases where aircraft maintenance was observed twice within the same week, within the same airport.

## **Spatiotemporal Semantic Knowledge Base**

The knowledge base (KB) is the repository for all data in the system. This includes data created by the analyst along with any inference from the ontology. The knowledge base therefore must support fast access using spatial extents, temporal extents, and combinations thereof. The knowledge base uses the Jena Semantic Web Framework [2] for query and graph processing, BBN's Asio Parliament KB [3] as an underlying RDF [4] storage mechanism, and libraries from BBN's Openmap GIS [5] application for spatial indexing.

Custom Jena Graph interfaces were developed to integrate Asio Parliament KB and the spatial and temporal indexes into the knowledge base. The custom interfaces encapsulate the implementation details, allowing transparent use by the query interface. The custom graph interface for the indexes facilitates ordering and splitting the queries between the semantic and spatiotemporal processing components.

## **Ontology Design**

The implemented ontology is designed to formalize a conceptual model of the world, enable a dynamic, context relevant user interface, and meet the data requirements of the overall system. The ontology is structured into three separate, but interrelated component ontologies. The foundational ontology formalizes the conceptual model used by the system and exists independent of the analytical domain. The domain ontology captures concepts of unique relevance to a domain (e.g. Air Defense). The application ontology meets the particular information requirements of the imagery analysis system. Each of these ontology components is discussed in more detail in the sections below.

## ***Foundational Ontology***

The foundational ontology is designed as an application independent, domain agnostic, conceptual model of the world,. It contains formalizations of basic notions of time and space and is a suitable model for information systems that maintain information about objects in the physical world over time. The foundational ontology is an integration and augmentation of best-of-breed, publicly available ontologies. The temporal representation used is OWL-Time [6,7], a product of the W3 Semantic Web Best Practices and Deployment Working Group. It includes interval and instant based time representations and is aligned with XML Schema built-in data types. This alignment eases application of existing RDF and XML software tools. The concrete geospatial representation is an adaptation of GeoRSS [8], which includes a profile of the Geography Markup Language (GML) [9]. Use of GML makes exchange of geospatial data with external tools feasible. OWL-Time and GeoRSS have both been integrated with the Basic Formal Ontology (BFO) [10]. BFO is a widely studied and published formal ontology that enumerates concepts at the highest levels of abstraction. In particular, all entities in the BFO formulation of the world are either continuants or occurrents. Continuants are those entities that have a continuous existence and endure through time. Examples include a piece of rock or the planet Earth. Occurrents are those entities that are bound in time and include processes and events. Examples of occurrents include walking the dog and the lifecycle of a frog.

## ***Domain Ontology***

The domain ontology used in this application formalizes air defense concepts. Many of these air defense concepts are adapted from publicly available sources of information on air defense topics, such as the Federation of American Scientists. The ontology is also aligned with National System for Geospatial-intelligence (NSG) feature catalog to promote reuse. This catalog provides a list of features and some relationships among the features. Names of features from the catalog are consistent with names used in the ontology. The NSG feature catalog does include subsumption (subclass/superclass relationships). These relationships are added, where appropriate, when NSG features are added to the domain ontology.

The domain ontology is aligned with the foundation ontology in order to determine which concepts are appropriate to populate the form-based UI for a given function. Specifically, some classes are subclasses of IndependentContinuant to express that they are standalone entities which an analyst can use to annotate an image (e.g. MiG-21). Other classes are subclasses of Qualities to indicate that these concepts can only be used as temporally changing attributes of an IndependentContinuant (e.g. the operational status of a MiG-21). Another example of alignment with the foundational ontology is that some classes are subclasses of Process. This indicates that these classes are to be used to indicate that some process or event is taking place (ex. fueling a MiG-21).

## ***Application Ontology***

The application ontology represents data that is specific to the function of this application. In other words, it contains application specific information. The imagery

analysis application ontology includes image metadata such as the date and time an image was taken and the name of the file.

## Results

The resulting environment provides an application through which a user can examine and annotate geo-registered imagery using air defense concepts and relationships described in the domain ontology. The inference capabilities provided by the ontology enable the system to automatically enrich each annotation and draw further conclusions. This allows users to search for annotations using abstractions and characteristics that were never specifically captured by the analyst. The spatiotemporal capabilities of the knowledge base combined with the semantics of the ontology enable analysts to efficiently query for observations that occur within a spatiotemporal extent or are related spatially or temporally. Finally, the representation of the ontology and data allows the annotations to be easily linked to annotations from other intelligence sources.

## Conclusion

This paper presents an imagery analysis environment that allows imagery to be annotated using highly descriptive semantic concepts and relationships defined in an ontology. By combining efficient semantic storage and retrieval techniques with efficient spatial and temporal indexing, these annotations can be queried and exploited in more powerful ways than can be achieved using traditional keyword search or relational database techniques.

## References

1. ESRI ArcGIS Desktop. *ESRI*. [Online] [http://www.esri.com/software/arcgis/about/desktop\\_gis.html](http://www.esri.com/software/arcgis/about/desktop_gis.html).
2. *Jena - A Semantic Web Framework for Java*. [Online] <http://jena.sourceforge.net>.
3. *Asio*. [Online] BBN Technologies. <http://www.asio.bbn.com>.
4. Resource Description Framework (RDF). [Online] <http://www.w3.org/RDF/>.
5. *OpenMap – Open Systems Mapping Technology*. [Online] <http://openmap.bbn.com>.
6. **Bechhofer, Sean, et al.** OWL Web Ontology Language Reference. *Worldwide Web Consortium*. [Online] February 10, 2004. <http://www.w3.org/TR/owl-ref/>.
7. **Hobbs, Jerry and Pan, Feng.** Time Ontology in OWL. *Worldwide Web Consortium*. [Online] September 26, 2006. <http://www.w3.org/TR/owl-time/>.
8. GeoRSS version 1.0. *GeoRSS::Geographically Encoded Objects for RSS feeds*. [Online] <http://georss.org/1>.
9. **Cox, Simon, et al.** *OpenGIS Geography Markup Language (GML) Implementation Specification, Version: 3.00*. s.l. : OpenGIS, 2003. Specification.
10. *SNAP and SPAN: Towards Dynamic Spatial Ontology*. **Grenon, Pierre and Smith, Barry.** 2004, *Spatial Cognition & Computation*, pp. 69-104.

# Ontological Support for Bayesian Evidence Management

**Michael N. Huhns and Marco G. Valtorta**  
**Center for Information Technology**  
**University of South Carolina**  
**Columbia, SC 29208**

## Abstract

This paper describes our work on an integrated system that can assist analysts in exploring hypotheses using Bayesian analysis of evidence from a variety of sources. The hypothesis exploration is aided by an ontology that represents domain knowledge, events, and causality for Bayesian reasoning, as well as models of information sources for evidential reasoning. We are validating the approach via a tool, Magellan, that uses Bayesian models for an analyst's prior and tacit knowledge about how evidence can be used to evaluate hypotheses.

## 1. Introduction

Much of the extensive work on ontologies to date has focused on modeling and representing the world of objects. The ontologies needed for our research supporting the management of hypotheses and evidence for analysts, however, must additionally model events and causality. Less work has been done on this aspect of ontologies. In this paper we show how concepts from a causal ontology can be used directly as variables in Bayesian networks and how the attributes of the causal concepts can be used in matching evidence to the variables. Moreover, subclass relationships in the ontology enable the extension of Bayesian reasoning over types.

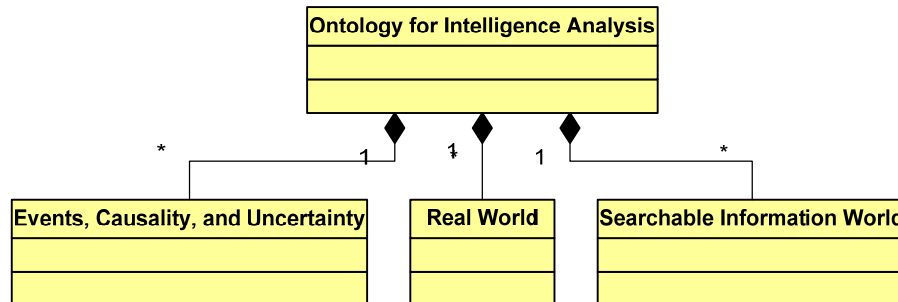
## 2. Bayesian Reasoning for Evidence Management

There are numerous real-world situations about which an analyst might wish to hypothesize and investigate, but it would be impractical to encode all of them explicitly in a support system for analysts. Instead, our approach is to represent fragments of situations and provide a mechanism for combining them into a wide variety of more complete ones [1,4]. The combination occurs dynamically as evidence about a situation becomes available or as an analyst revises or enters new hypotheses. A fragment is represented as a Bayesian network with nodes for hypotheses, events, and evidence, and links for relating them. Our ability to combine the fragments into more complete situation models is dependent on having a consistent terminology in which the fragments are described. The focus of our work has been on (1) defining and representing the terminology, including terms of a domain and terms for evidence in that domain, (2) capturing new fragments from a variety of sources, and (3) incorporating the terminology and BN fragments into an integrated end-to-end tool, Magellan.

### 2.1. *Capturing the Terminology and Prior Knowledge for a New Domain*

Intelligence analysts are concerned primarily with hypotheses that involve cause-and-effect. These are best supported by an ontology emphasizing events and their causal relationships, along with a hypothetical world of possible events, actions, and causes. However, causal relationships must be interpreted in the context of real-world objects and their properties, which can be represented in a conventional ontology such as those that are part of SUMO. The evidence for reasoning about hypotheses can come from a variety of sources, and the acquisition of evidence and events from these sources must also be represented, constituting a third kind of ontological representation describing the information sources. Figure 1 depicts the three ontological models we use for modeling situations, relating situations to background knowledge about objects in the

world, acquiring evidence, and assessing the likelihood of the situations using Bayesian reasoning. Our tool, Magellan, uses Protégé for capturing the ontologies, RDF for representing the terminology, XMLBIF for representing the causal relationships, and RDF and RDQL for requesting evidence from information sources.



**Figure 1. An ontology for intelligence analysts has three related parts, corresponding to the world of causality and hypothetical events needed for Bayesian reasoning, the real world of things needed to model situations, and the world of information and information sources needed for evidence management**

Causality is a special relationship among events for which certain properties hold probabilistically. For example, causality is logically irreflexive and asymmetric, but probabilistically transitive. Causal models are very useful, because they allow prediction of the effect of interventions [3,5].

New variables are added to the causal and event portion of an analyst’s ontology using Protégé, so that all of the nodes in a Bayesian network fragment are represented in a standard and consistent terminology. We extend SUMO with this terminology, so that we can take advantage of SUMO’s existing description of general knowledge of the world. Each variable has a set of identifying attributes, which are used to combine fragments (fragments can be combined only if their attributes unify) [4].

Probabilities are assigned to events in the fragment by performing experiments, estimating beliefs, or counting outcomes. Once assigned, they are updated by conditioning on evidence using Bayes rule and the laws of probability. The fragments are stored in a repository, where they can be matched with evidence and combined with other fragments to produce models of situations that are as complete, accurate, and specific as possible.

We also represent in the information source ontology the level of credibility of items of evidence, and provide a Bayesian interpretation of credibility. We define *evidence* to be a collection of findings, each of which describes the state of a Bayesian network variable, and distinguish three kinds [7]:

1. A *hard finding* specifies that the variable has a particular value.
2. A *soft finding* is a distribution on the states of a variable, usually corresponding to an “objective” statistical distribution that is not expected to change within a scenario.
3. A *virtual finding* is a likelihood ratio corresponding to the credibility associated to an evidence source, such as a witness. Unlike soft findings, virtual findings allow for an update of the posterior probability of the evidence variable.

Our modified version of ACH1 [6] is used by an analyst to enter the appropriate hypotheses and any initial evidence that might be available. The terminology available to the analyst is provided via drop-down menus as shown in Figure 2, where the menu entries are the ontology terms from



Protégé. The resultant ACH [2] matrix is converted automatically into a bipartite Bayesian network, with initial probabilities assigned based on the relevance factors assigned to cells of the matrix. The network is saved into the repository of fragments.

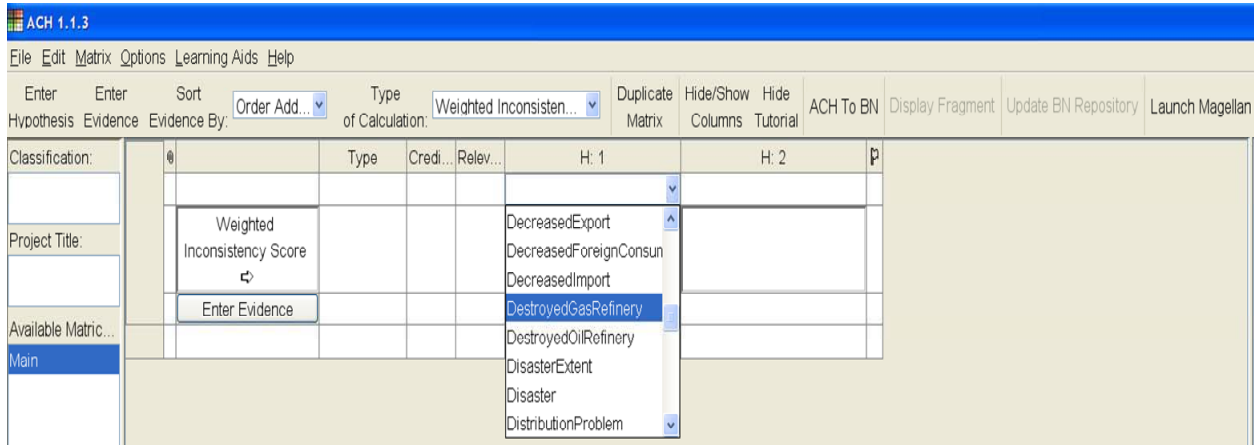


Figure 2. The extended ACH interface is integrated with the ontology of events through pull-down menus

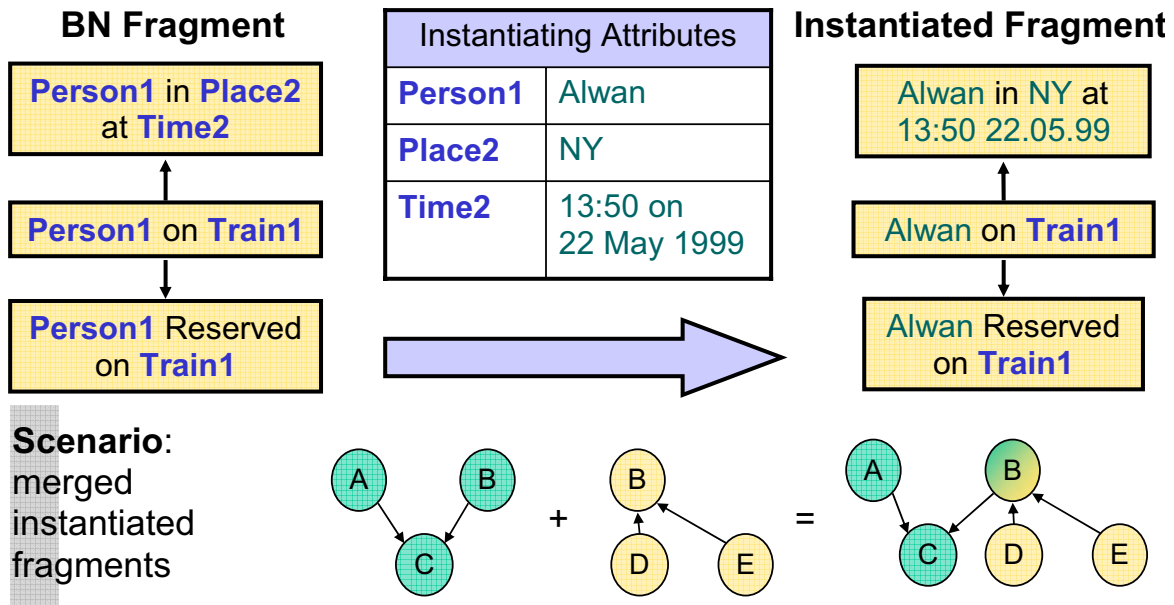


Figure 3. Fragments (templates) are merged based on instantiating evidence

### 3. Architecture for Bayesian Reasoning

Figure 4 shows an architecture for Bayesian reasoning, which would be used as follows. Based on initial triggering messages, or based on a hypothesized situation that an analyst would like to investigate, an appropriate scenario represented as a Bayesian model is chosen by the analyst and a corresponding form is shown listing initial evidence and the domain variables for the scenario. The evidence values for the variables can be supplied automatically from the triggering messages or can be entered by the analyst. The Bayesian reasoning component, using a value-of-information calculation, then determines which pieces of evidence would be most useful in confirming or denying the analyst's hypothesis. A request for this evidence is sent to the analyst, who returns the result to the Bayesian reasoner for incorporation into the scenario, and the

likelihood of the analyst’s hypothesis is reassessed. The process is repeated until the analyst decides to stop or there is no more evidence available that changes the plausible outcomes.

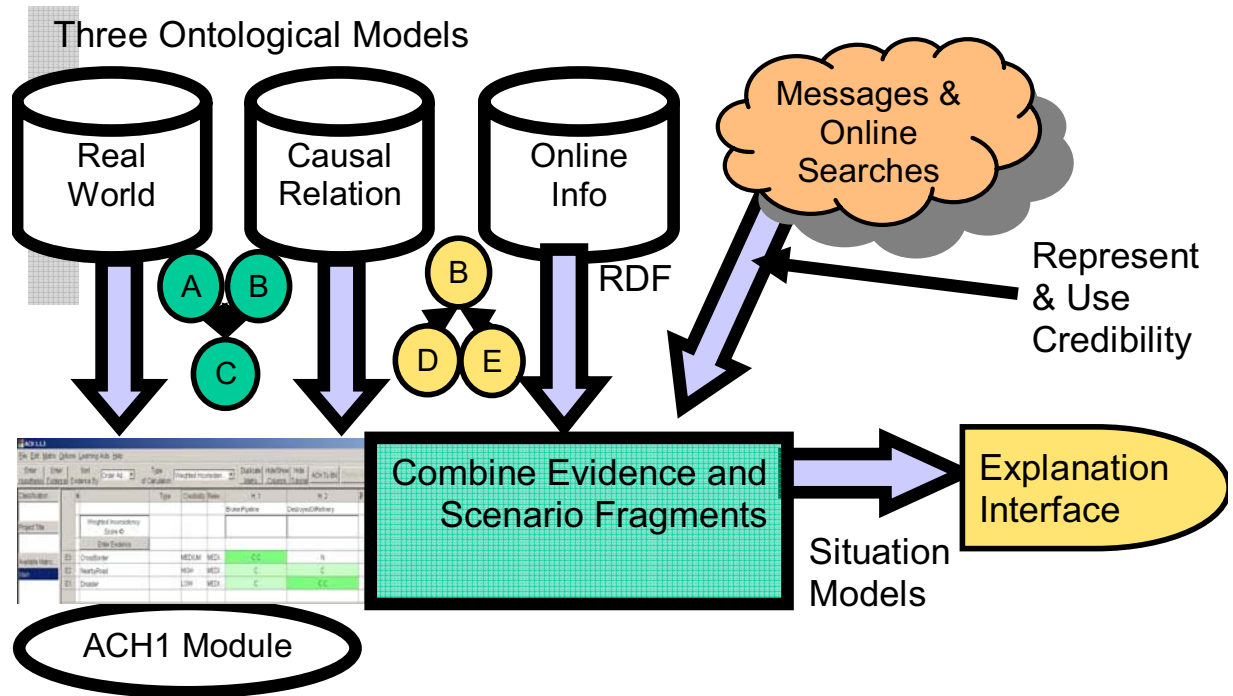


Figure 4. Magellan architecture for Bayesian Reasoning used to explore an analyst’s hypotheses

#### 4. Conclusion

As we continue to increase the functionality of our Bayesian reasoning system, we will improve our representation of events and causality, and increase the capabilities for the application of prior and tacit knowledge to the exploration of analysts’ hypotheses.

#### 5. Acknowledgements

This work was funded in part by the Disruptive Technology Office Collaboration and Analyst System Effectiveness (CASE) Program, contract FA8750-06-C-0194 issued by Air Force Research Laboratory (AFRL). The views and conclusions are those of the authors, not of the US Government or its agencies.

#### 6. References

- [1] John Cheng, Ray Emami, Larry Kerschberg, Eugene Santos, Jr., Qunhua Zhao, Hien Nguyen, Hua Wang, Michael Huhns, Marco Valtorta, Jiangbo Dang, Hrishikesh Goradia, Jingshan Huang, and Sharon Xi, “OmniSeer: A Cognitive Framework for User Modeling, Reuse of Prior and Tacit Knowledge, and Collaborative Knowledge Services,” *Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS38)*, Hawaii, January, 2005.
- [2] Richards J. Heuer, Jr., *Psychology of Intelligence Analysis*, Center for the Study of Intelligence (at <http://www.cia.gov/csi/books/19104/index.html>), 1999.
- [3] Yimin Huang and Marco Valtorta, “Pearl’s Calculus of Intervention is Complete,” *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI-06)*, pp. 217-224, 2006.

- [4] Katherine Laskey and Suzanne Mahoney, "Network Fragments: Representing Knowledge for Constructing Probabilistic Models," *Proceeding of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, pp. 334-341, 1997.
- [5] Judea Pearl, *Causality: Modeling, Reasoning, and Inference*, Cambridge, England: Cambridge University Press, 2000.
- [6] Peter Pirolli and Lance Good, "Evaluation of a Computer Support Tool for Analysis of Competing Hypotheses," Palo Alto, CA: Palo Alto Research Center (UIR Technical Report), 2004.
- [7] Marco G. Valtorta, Y.-G. Kim, and Jirka Vomlel, "Soft Evidential Update for Multiagent Systems," *International Journal of Approximate Reasoning*, 29, 1, pp. 71-106, January 2002.



## Achieving Ontology-Assisted Query of Graph Databases

David Silberberg, Wayne Bethea, Dennis Patrone, Paul Frank,  
David Patrone, John Gersh, Elisabeth Immer  
Johns Hopkins University Applied Physics Laboratory  
11100 Johns Hopkins Road  
Laurel, MD 20723-6099

### Abstract

We describe an approach for enabling ontology-assisted queries onto existing schema-based graph database systems without altering the graph query language or the corresponding graph database system. Typical schema-based graph database systems enable analysts to formulate queries using terms from a schema. Our approach enables analysts to formulate queries using terms from a *virtual schema*, which is composed of an ontology, a graph schema, and mappings between them. A software system can then assist the analyst by extracting the predicates and terms from queries, and in conjunction with the ontology and a reasoner, produce a set of corresponding graph queries that contain only terms from the graph schema. These queries are then sent to the graph database for evaluation. This approach enables intelligence analysts to focus on analysis that is more complex while the ontology-assisted query capability performs lower level reasoning. A distinction is maintained between the ontology reasoning and graph query systems to 1) take advantage of the performance of graph query engines while exploiting the semantics of the ontologies, 2) provide multiple analysts with an explicit and consistent semantic model of the graph data, and 3) enable multiple analysts with different semantic models of the data to use their own personal ontologies for analysis.

### PLEASE NOTE:

**PERMISSION TO PUBLISH THIS PAPER HAS NOT BEEN  
GRANTED AS OF THE TIME OF PRINTING OF PROCEEDINGS.**



## Geospatial Ontology Trade Study

James Ressler<sup>1</sup>, Mike Dean<sup>2</sup>

<sup>1</sup>[james.ressler@ngc.com](mailto:james.ressler@ngc.com) Northrop Grumman Corporation  
1010 Market Street, Suite 1740, St. Louis, Missouri, 63101

<sup>2</sup>[mdean@bbn.com](mailto:mdean@bbn.com), BBN Technologies  
1300 N. 17<sup>th</sup> Street, Suite 400, Arlington, Virginia, 22209

**Abstract.** This short paper summarizes a survey of ontologies relevant to geospatial intelligence. 45 geospatial and temporal ontologies, in 11 categories, were assessed against 3 use cases: annotation, qualitative reasoning, and information integration. Specific recommendations and more general conclusions are provided. The paper presents an illustration of a feature with several different ontology representations.

**Keywords:** ontology, geospatial intelligence, Semantic Web, OWL, temporal

### Introduction

An ontology is a formal, explicit, shared conceptualization of a domain. As such, it defines the concepts and vocabulary used within a community of interest. The formal, machine-understandable representation allows use of the ontology to support logical inference. The Semantic Web has motivated increased development and use of ontologies, while at the same time driving the need for common ontologies.

A trade study conducted a broad survey of 45 ontologies that apply to the spatial, event, and temporal granularity concepts [3]. The study has reviewed thousands of classes and properties in order to find the characteristics that make these ontologies best suited to the geospatial intelligence community for uses in annotation, qualitative reasoning and interoperability. The study has made several strategic recommendations to further the development of semantic technology and the application of these ontologies.

To focus discussion and applicability, the study considered three primary use cases motivating the development and selection of geospatial and related ontologies.

- **Annotation:** using classes and properties to represent relevant characteristics of objects
- **Qualitative Reasoning:** reasoning about spatiotemporal relationships between objects (e.g. containedWithin, connectedTo, During)
- **Information Integration:** facilitate interoperability by mapping other data models to/from a common encompassing reference ontology

The study used seven categories of Geospatial Ontologies<sup>1</sup> from the W3C Geospatial Incubator Group with four additional categories to fully represent the range of geospatial and temporal data:

---

<sup>1</sup> <http://www.w3.org/2005/Incubator/geo/>

1. Geospatial Feature Ontology
2. Feature Type Ontology
3. Spatial Relationship Ontology
4. Toponym (Place name) Ontology
5. Coordinate Reference / Spatial Grid Ontology
6. Geospatial Metadata Ontology
7. (Geospatial) Web Services Ontology
8. Geometric Ontology
9. Coverage Ontology
10. Geopolitical Ontology
11. Temporal Ontology

## Ontologies Surveyed

The study included the following ontologies:

- Basic Formal Ontology
- DOLCE
- Cyc (Geodesy, Linear Object, Map Projection, Open Geospatial Consortium, Surface Geometry, Temporal Predicates, Time Interval, Terrain, and Topology domains)
- FGDC Content Standard for Digital Geospatial Metadata
- ISO Geographic Information (Conceptual Schema Language/19103, Spatial/19107, Temporal/19108, Application Schema/19109, Feature Cataloguing/19110, Spatial Reference-coordinates/19111, Geographic Identifier/19112, Metadata/19115, Coverage/19123)
- SOUPA (RCC, Geomeasurement, Event, Time)
- SUMO (SUMO, MILO, Geography)
- Enterprise Conceptual Data Model
- NSG Application Schema
- georSS (BasicGeo, NEOGEO)
- Geography Markup Language (ISO 19136)
- Keyhole Markup Language
- geonames.org
- MINDSWAP (geoCoordinateSystems, geoFeatures, geoRelations)
- SWEET (Space, Time)
- S-57 (maritime domain)
- OWL-Time
- RDF Calendar

Representations of most of the ontologies were available in the W3C OWL Web Ontology Language [2], which was the focus of the study. Some were originally



developed using other representations and then converted to OWL, sometimes with some loss of expressivity. Several had not yet been represented in OWL.

## Conclusions

The principal recommendations of the study were:

- There are a number of acceptable OWL ontologies and related representations available that can be re-used and extended for a domain.
- Update, formalize and control ontologies in a best practice document aligned with existing standards.
- Create an ontology library for the ISO Geographic Information technical committee specifications and encourage the U.S. to submit these ontologies to ISO for approval.
- Use the National System for Geospatial Intelligence (NSG) Application Schema as the basis of a standard feature type ontology.
- Representatives from the geospatial intelligence community should participate in the new ISO Technical Committee 211 project 19150 to promote spatial upper ontologies.

The study also recommends the following guidelines for reusing geospatial ontologies:

- Use OWL for ontology definition.
- Use the simplest OWL representations that meet application needs.
- Geospatial Ontologies should be based upon standards consistent with the NSG Architecture and with the GEOINT Standards listed in the Defense Information Standards Registry (DISR), which are also contained in the NSG Architecture Compliance.

For the Geospatial Intelligence user, a recommended ontology/set of ontologies for each of the ontology categories and use cases described is shown in Table 1. Each row and column also includes the number of applicable ontologies. For each category or use case, these criteria were used: “Fully” means that the concepts in the ontology directly apply without modification; “Partially” means that some of the concepts apply, but it is not the primary intent of the ontology or requires modification; “Indirectly” means that while not being directly applicable, the ontology contributes towards application. Several categories are marked “Not Applicable” because the ontology category was not intended for the use case (e.g. Qualitative Reasoning). The rationale for each category recommendation is contained in the trade study report [3].

Ontology Category	Recommended Ontology per Use Case		
	Annotation (15 fully, 12 partially, 5 indirectly)	Qualitative Reasoning (7 fully, 8 partially, 2 indirectly)	Information Integration (8 fully, 7 partially, 21 indirectly)
1. Geospatial Feature (7 fully, 6 partially)	GeoRSS (Simple or GML)	Not applicable	GML
2. Feature Type (6 fully, 2 partially)	NSG FC / NAS 1.8 (in OWL)	Not applicable	NSG FC / NAS 1.8 (in OWL)
3. Spatial Relationship (3 fully, 6 partially)	SOUPA rcc	SOUPA rcc	SOUPA rcc
4. Toponym (1 fully, 4 partially)	ISO 19112	Not applicable	ISO 19112
5. Coordinate Reference (4 fully, 2 partially)	ISO 19111, Cyc Map Projection	Not applicable	ISO 19111
6. Geospatial Metadata (2 fully, 3 partially)	ISO 19115	Not applicable	ISO 19115
7. Web Service (0 fully, 2 partially)	Not evaluated	Not evaluated	Not evaluated
8. Geometric (4 fully, 4 partially)	ISO 19107	SOUPA rcc	ISO 19107
9. Coverage (1 fully, 4 partially)	None available	Not applicable	None available
10. Geopolitical (2 fully, 3 partially)	None recommended	Not applicable	None recommended
11. Temporal (6 fully, 4 partially)	XML datatypes in OWL, OWL-Time	OWL-Time	OWL-Time

**Table 1: Recommended Spatiotemporal Ontologies**

In addition to these recommendations, these follow-up actions were recommended.

- Existing ontologies that have unique and useful concepts, such as Cyc, MINDSWAP, and SUMO, should be linked to NSG ontologies and augmented with NGA specific domain concepts from the ECDM and GSIP.
- Perform a metrics evaluation of the quality of the selected ontologies similar to the assessment performed by Burton-Jones on the DAML ontology library [1]. This quality assessment developed measurements of an ontology's syntax, richness, interpretability, clarity, comprehensibility and relevance.
- Given the active interest in service oriented architectures, the use of ontologies to describe services (such as OWL-S and SAWSDL) is an active area of research and commercial development. An evaluation of ontologies to represent web services is recommended for a future study.

The Spatial Ontology Community of Practice (SOCoP) of the US Federal CIO Council provides a good forum for exposing and coordinating geospatial ontologies. As intelligence agencies employ semantic technology, interoperability should be considered

from the outset because semantic queries are not inherently interoperable when performed across domains. It is in the best interest of the Intelligence Community to act on these recommendations and guidelines to provide interoperable and mature semantic technology.

## References

[1] Burton-Jones, A., Storey, V., Sugumaran, V., Ahluwalia, P. “A semiotic metrics suite for assessing the quality of ontologies.” *Data & Knowledge Engineering* 55 (2005) 84-102.

[2] M. Dean and G. Schreiber, eds. OWL Web Ontology Language Reference. W3C Recommendation 10 February 2004. <http://www.w3.org/TR/2004/REC-owl-ref-20040210/>

[3] Ressler, J., Dean, M. and Kolas, D., “Geospatial Ontology Trade Study”, [http://projects.semwebcentral.org/docman/?group\\_id=84](http://projects.semwebcentral.org/docman/?group_id=84), June 2007..

## Appendix: Illustration of Feature Ontology

To ground the analysis and allow comparisons across ontologies, the study included annotations for a specific feature: The Pentagon in Washington, D.C., which is in the shape of a five-sided polygon as illustrated in Figure 1.



**Figure 1: Pentagon Feature (GoogleEarth © Google)**

The representation of The Pentagon feature in GeoRSS Simple is as follows:

```
<rdf:Description>
  <georss:featurename>Pentagon</georss:featurename>
  <georss:polygon>
    -77.05795370823761 38.87258672915797
    -77.05847549720639 38.87005708744568
    -77.0555999760046 38.86886750371786
    -77.05326581781736 38.87064560343153
    -77.05465594662488 38.87292421787603
    -77.05795370823761 38.87258672915797
  </georss:polygon>
</rdf:Description>
```

GeoRSS GML can encapsulate the GML representation of the feature:

```
<gml:FeatureCollection
  xmlns:gml="http://www.opengis.net/gml"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.safe.com/gml/fme
pentagon3.xsd">
  <gml:boundedBy>
    <gml:Envelope srsName="EPSG:4326" srsDimension="2">
      <gml:lowerCorner>-77.0584754972064
38.8688675037179</gml:lowerCorner>
      <gml:upperCorner>-77.0532658178174
```

```

38.872924217876</gml:upperCorner>
  </gml:Envelope>
</gml:boundedBy>
<gml:featureMember>
  <gml:surfaceProperty>
    <gml:Surface srsName="EPSG:4326" srsDimension="2">
      <gml:patches>
        <gml:PolygonPatch>
          <gml:exterior>
            <gml:LinearRing>
              <gml:posList>
                -77.0579537082376 38.872586729158
                -77.0584754972064 38.8700570874457
                -77.0555999760046 38.8688675037179
                -77.0532658178174 38.8706456034315
                -77.0546559466249 38.872924217876
                -77.0579537082376 38.872586729158
              </gml:posList>
            </gml:LinearRing>
          </gml:exterior>
        </gml:PolygonPatch>
      </gml:patches>
    </gml:Surface>
  </gml:surfaceProperty>
</gml:featureMember>
</gml:FeatureCollection>

```

The representation of the sample Pentagon feature in KML is as follows:

```

<Placemark>
  <name>The Pentagon</name>
  <styleUrl>#msn_ylw-pushpin</styleUrl>
  <Polygon>
    <tessellate>1</tessellate>
    <outerBoundaryIs>
      <LinearRing>
        <coordinates>
          -77.05795370823761,38.87258672915797,0
          -77.05847549720639,38.87005708744568,0
          -77.0555999760046,38.86886750371786,0
          -77.05326581781736,38.87064560343153,0
          -77.05465594662488,38.87292421787603,0
          -77.05795370823761,38.87258672915797,0
        </coordinates>
      </LinearRing>
    </outerBoundaryIs>
  </Polygon>
</Placemark>

```

Geonames.org returns the following information for The Pentagon:

```
<Feature rdf:about="http://sws.geonames.org/4778469/">
  <name>Pentagon</name>
  ...
  <alternateName xml:lang="fr">Pentagone</alternateName>
  ...
  <featureClass rdf:resource="http://www.geonames.org/ontology#S"/>
  <featureCode rdf:resource="http://www.geonames.org/ontology#S.BLDG"/>
  <inCountry rdf:resource="http://www.geonames.org/countries/#US"/>
  <wgs84_pos:lat>38.8709455</wgs84_pos:lat>
  <wgs84_pos:long>-77.0552551</wgs84_pos:long>
  <parentFeature rdf:resource="http://sws.geonames.org/4744725"/>
  <nearbyFeatures
    rdf:resource="http://sws.geonames.org/4778469/nearby.rdf"/>
  <locationMap>http://www.geonames.org/4778469/pentagon.html</locationMap>
  <wikipediaArticle
    rdf:resource="http://de.wikipedia.org/wiki/Pentagon"/>
</Feature>
```

The representation of the Pentagon example using SWEET is as follows:

```
<material_thing:Building>
  <space:hasBoundary>
    <numerics:Polygon>
      <numerics:hasVertices>
        <numerics:Point>
          <numerics:hasCoordinates>
            <space:GeographicalCoordinates>
              <rdf:_1>
                <space:Longitude>
                  <numerics:hasValue>77.05795370823761</numerics:hasValue>
                </space:Longitude>
              </rdf:_1>
              <rdf:_2>
                <space:Latitude>
                  <numerics:hasValue>38.87258672915797</numerics:hasValue>
                </space:Latitude>
              </rdf:_2>
            </space:GeographicalCoordinates>
          </numerics:hasCoordinates>
        </numerics:Point>
      </numerics:hasVertices>
      <!-- ... -->
    </numerics:Polygon>
  </space:hasBoundary>
</material_thing:Building>
```

# A PRAGMATIC FOUNDATION FOR DEFINING A RICH SEMANTIC MODEL OF *TRACK*

RICK HAYES-ROTH and CURTIS BLAIS

*Naval Postgraduate School  
Monterey, California USA*

## INTRODUCTION

Many defense, homeland security, and commercial security objectives require continuous tracking of mobile entities. The systems that perform these functions produce information products called *tracks*. A track associates observations with the mobile entity and typically includes position, velocity, and other similar attributes. Military systems have sophisticated tracking and track fusion processes, but lack uniformity in syntactic and semantic content preventing effective sharing of the information. In other domains of interest, such as seagoing surface ships, dangerous cargo and persons of interest, tracking systems are less mature and have marginal performance. It is now essential that we be able to share information across different tracking systems working in related domains.

In this paper, we describe the Rich Semantic *Track* model [Hayes-Roth 2005] as a foundation for sharing world state information across multiple systems. The model exhibits a belief and evidentiary structure that has not been emphasized in previous track models for broad application. The approach is having a significant impact on design of emerging models, particularly the Maritime Information Exchange Model.

## THE PRAGMATICS OF *TRACK*

*Tracks* are an important element of situation assessment in most command and control systems. Commanders want to track platforms and forces, anticipate their likely motions and potential threats, determine how best to counter threats, and then implement chosen countermeasures efficiently. From these general concerns, we identify the following common pragmatic objectives for a mobile entity  $M$  with possible intentions and capabilities to do harm to our interests:

- (1) Observe, detect, identify, classify and continuously monitor  $M$ .
- (2) Locate  $M$ .
- (3) Infer  $M$ 's intent.
- (4) Determine  $M$ 's threats  $T_{M,D}$  against domain  $D$ .
- (5) Predict  $M$ 's future location and behavior.
- (6) Alert agent  $A$  about  $M$  and threats  $T_{M,D}$ .
- (7) Determine countermeasures  $CM(T_{M,D})$  to threats  $T_{M,D}$ .
- (8) Inform agent  $A$  about countermeasures  $CM(T_{M,D})$ .

These eight pragmatic objectives define the general and common concerns of military and security agencies with potentially dangerous mobile entities. The whole purpose of sharing information among different sources is to support these common objectives.

Any system of concepts will have its own nuances and best practices for modeling the world effectively. No system is perfect; instead, we wish to initiate use of **evolvable semantics to**

**support important pragmatics.** Thus, the key capability we need is to do some things well while being able to improve continually. For that reason, almost any reasonable semantic system will be good enough for significant information sharing. The essential quality required is that the system distinguishes states that warrant different inferences and actions.

All assertions in the information space about the state of the world (such as about vessels, cargo, people) are *beliefs*. So, every aspect of the information model of tracks should be considered a “belief” with whatever supporting data any belief can have. Here are the most common structures:

- (1) A belief is held to be a **fact**.
- (2) A belief is a widely accepted **assumption** that’s recognized to be less certain than a fact.
- (3) A belief is based on direct credible eye witness report, so it’s like **ground truth**.
- (4) A belief is based on summarizing and aggregating other beliefs so it’s a logical inference or **implication**.
- (5) A belief is based on the association and fusion of K observations that support a simplifying inductive inference, interpretation or **abduction**.
- (6) A belief is a **composition** (AND) of other beliefs.
- (7) A belief is a probable inference or **confirming prediction** from another belief.
- (8) A belief is an improbable inference from another belief or a **disconfirming expectation**.
- (9) A belief is an analyst **judgment**, intuition, opinion, or concern, based on some other beliefs as well as some inference.
- (10) A belief is a pattern-based or rule-based **assessment**, where a set of beliefs about an entity instantiates a pattern template above some threshold level indicating that the pattern’s interpretation applies.

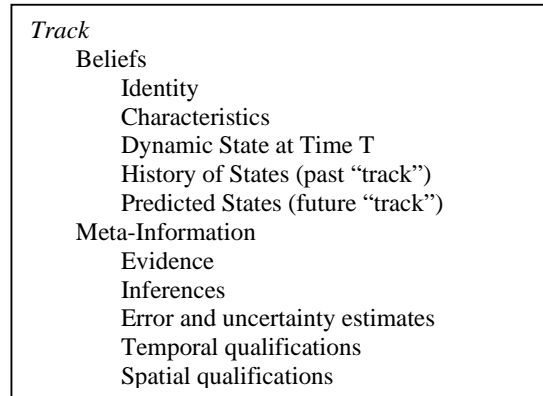
Therefore, our approach is to identify a rich semantic model of *tracks* that can express these fundamental belief structures in order to: represent a wide variety of meanings and support a broad array of pragmatic goals; reduce implementation time and cost required to reason about a new type of *track*; simplify the understanding and importation of external sources of track information; help operators describe track attributes they value in performing their tasks; improve our ability to combine multiple sources of track information; provide a stable and evolvable base for best practices supporting information sharing; and improve bandwidth utilization by delivering nothing but valued information at the right time (VIRT) [Hayes-Roth, 2004, 2006].

## **THE SEMANTICS OF *TRACK***

The choice of engineered semantics rests on pragmatics – describing what differences in behavior must be supported. Given a set of pragmatic objectives, the inference process considered earlier relies upon conceptual categories. A semantic hub or “core” should make all of the conceptual distinctions required to support those categories and related pragmatics. The rich semantic *Track* model, therefore, should reflect aspects of state that most users of track information require for addressing expected pragmatic concerns. As we employ such a model to mediate sharing among systems, we will inevitably discover additional concerns not yet adequately addressed in the current model. This will drive an iterative, evolutionary series of improvements to the community’s evolving model of *Track*.



We have created a mostly-hierarchical conceptual scheme (Figure 2) working backwards from pragmatic objectives to required concepts to supporting distinguished data values. The ability to adapt this standard hierarchy rapidly to exploit a new source would be the operational test of value. This suggests both what types of products we need and also what types of methods will enable us to adapt these products to new situations. The *Track* model allows us to describe our beliefs about a mobile entity and its past, present and predicted future states. In addition, we are able to justify inferences that we make as part of the tracking process.



**Figure 2. The top-level conceptual hierarchy for *Track*<sup>1</sup>.**

This fragment of a conceptual hierarchy describes the most general, or topmost, element called *Track*. The concept *Track* contains two principal component concepts, called Beliefs and Meta-Information, respectively. Components of Meta-Information may apply to each element of Beliefs. That is, when we use the conceptual hierarchy to create actual beliefs that are instances of *Track* Beliefs, we may find it useful to qualify every belief by using the sub-concepts of Meta-Information. In this sense, Meta-Information plays dual roles of meta-data (data about data) or reification (statements about statements). Moreover, Meta-Information can apply to combined Beliefs, as in providing rationale for bringing the Beliefs together.

## **RESEARCH AND DEVELOPMENT AGENDA TO ACHIEVE POTENTIAL BENEFITS OF THE SEMANTIC MODEL OF *TRACK***

To advance the agenda on track-related systems, we need to accomplish several intermediate objectives:

- Select a community of interest that recognizes the importance of this task.
- Enumerate and prioritize information sharing scenarios.
- Determine a high-value near-term subset of the hub semantics.
- Identify viewer/editors that operators will employ in these sharing scenarios.
- Determine translator requirements to support the scenario.
- Implement an initial semantic hub and related translators to/from interoperating systems..
- Test the environment, and identify high priority requirements for improvements to the hub and translators.
- Identify operators for whom VIRT capabilities have highest value.
- Determine best methods to gain knowledge of operator’s context and identify valuable information.

---

<sup>1</sup> Successively indented topics represent specializations or subcategories under the topic they descend from.

- Implement query methods and notification methods to operationalize valued information at the right time.
- Iterate, through earlier steps, to implement continuous improvement.
- Place responsibility for this continuous improvement process in the hands of an appropriate agent or team.

This R&D agenda provides an incremental approach that can provide immediate benefits and can quickly exploit learning to gain additional benefits. These concepts are already influencing new track model designs. The approach has informed development of the Joint Track Management data model and is strongly integrated into the Comprehensive Maritime Awareness (CMA JCTD) Maritime Information Exchange Model (MIEM). In the MIEM, all objects and their constituent elements support a rich metadata structure (information sources, pedigree, time-varying nature, threat/vulnerability, confidence, annotation, etc.) to enable clear expression of value added information.

## CONCLUSIONS

Many defense, homeland security, and commercial security objectives require continuous tracking of mobile entities. We wish to share information among different tracking systems working in similar domains. To combine information from different sources, we will need a flexible framework that can tolerate and exploit data products from different systems, although these systems employ different representations and embody different assumptions. Our approach is to create a rich semantic model of tracks that can support a wide variety of objectives related to information sharing. The semantic model is developed to play the role of a hub amidst a variety of translators. This approach enables achievement of significant positive benefits through incremental improvement driven by pragmatic concerns.

## REFERENCES

- Hayes-Roth F. 2004. "Model-based communication networks for improved collaboration and decision-making," presented at the *12th International Conference on Telecommunication Systems: Modeling and Analysis*, Naval Postgraduate School, Monterey, 2004.  
<http://doncio.ro.nps.navy.mil/icts12/>
- Hayes-Roth F. 2005. "Towards a Rich Semantic Model of Track: Essential Foundation for Information Sharing". Research Paper, Naval Postgraduate School, Monterey, CA. February 25, 2005.
- Hayes-Roth F. 2006. "Model-Based Communication Networks and VIRT: Orders of Magnitude Better for Information Superiority". In *Proceedings of the Military Communications Conference* (Washington, DC, October).

# Toward Automated Provability-Based Semantic Interoperability Between Ontologies for the Intelligence Community

(extended abstract)

Andrew Shilliday, Joshua Taylor, Selmer Bringsjord

{shilla,tayloj,selmer}@rpi.edu

Department of Cognitive Science

Department of Computer Science

Rensselaer AI & Reasoning (RAIR) Lab:

<http://www.cogsci.rpi.edu/research/rair/>

Troy NY 12180 USA

September 27, 2007

## 1 Introduction

The need for interoperability is dire: Knowledge representation systems employ ontologies that use disparate formalisms to describe related domains; to be truly useful to the intelligence community, they must meaningfully share information. Ongoing research [3, 4, 7, 15] strives toward the holy grail of complete interoperability, but has been hindered by techniques that are specialized for particular ontologies, and that lack the expressivity needed to describe complex ontological relationships. In the sequel, we describe *provability-based semantic interoperability* (PBSI) [16], a means to surmount these hindrances; *translation graphs*, one of our key formalism for describing the complex relationships among arbitrary ontologies; and ways in which these techniques might be automated.

## 2 PBSI and PBSI<sup>+</sup>

We clarify our uses of *syntactic* and *semantic*. The *syntax* of a knowledgebase regiments the structure of expressions in it (e.g., that `(mother-of Amy)` is a well-formed KIF term owes to KIF's syntax); *semantics* attribute *meaning* to otherwise abstract constructs (`(mother-of Amy)` designates Amy's mother according to the semantics of an ontology). A *syntactic translation* occurs

when knowledge from one ontology is moved into another using the same semantics. In other words, when ontologies describe the same *kind* of things, and differ only in the way object-level information is structured, interoperability is achieved by mere syntactic translation. When ontologies differ not only in syntax, but also in semantics (yet relate meaningfully), a stronger form of translation is needed: *semantic translation* enables the transfer of information across such ontologies. Systems capable of semantic translation (e.g., [4, 6]) provide some language in which to formalize the semantic connections between ontologies. Unfortunately, the relationships associating ontologies may be so complex that translation of knowledge from one ontology into another is not feasible. Moreover, when interoperability is achieved between complex ontologies, justification is needed to support trust that the meaning of the data has been preserved.

PBSI provides a language for formalizing the relationships between ontologies via *bridging axioms*, and our extension, PBSI<sup>+</sup>, associates each information exchange with a proof certifying the conservation of semantic meaning. The basic construct of PBSI<sup>+</sup> is the signature, a collection of statements in the *meta-theory* which, coupled with a set of axioms, captures a given ontology. A signature  $\Sigma$  consists of a set  $\sigma$  of sorts, and a set  $\phi$  of functors. A sort  $s \in \sigma$  is a domain — a collection whose elements are considered the same *kind* of thing,<sup>1</sup> (e.g., the months in the year, boolean values, natural numbers, US citizens). A functor  $f \in \phi$  maps between objects of the sorts in  $\sigma$ . In the case that  $f$  maps onto the boolean values,  $f$  is a relation; if it also takes no arguments, it is a proposition. Having defined signatures, the specifications of ontologies, we present *translation graphs*, a framework for bridging signatures (and so, ontologies) while preserving semantics.

### 3 Translation Graphs

A translation graph, like the one in figure 1, is a directed graph  $G = (V, E)$  where the vertices  $v \in V$  are each unique signatures, and each edge  $e = (u, v) \in E$  describes the application of a primitive operation to  $u$  yielding  $v$ , viz., adding or removing either a sort or functor. The addition of a new functor also has associated information potentially relating the new functor to existing functors of the modified signature.

As a toy example, let signature  $\Sigma_1$  consist of the domains  $\sigma_1 = \{\text{People}, \text{Firearms}\}$  and just one functor  $\phi_1 = \{\text{OwnerOf} : \text{Firearms} \rightarrow \text{People}\}$ , which is understood to map a firearm to its owner. Furthermore, signature  $\Sigma_2$  consists of the domain  $\sigma_2 = \{\text{People}\}$  and the functor  $\phi_2 = \{\text{IsArmed} : \text{People} \rightarrow \text{Boolean}\}$  so that `IsArmed` holds for those people who own guns (in this example, all signatures implicitly have the boolean domain). A translation graph enabling interoperability between these signatures might apply the following primitive operations bridging  $\Sigma_1$  to  $\Sigma_2$ :

1. *AddFunctor*(`IsArmed`) with the bridging axiom

$$\forall_p [\exists_g \text{OwnerOf}(g) = p] \rightarrow \text{IsArmed}(p)$$

so that the the relation `IsArmed` holds for any person,  $p$  where there is a firearm that  $p$  owns.

---

<sup>1</sup>Our current formalization draws on many-sorted logic, and so domains are disjoint. While this is a limitation on the expressivity of the language (many ontologies require a subsort hierarchy), it is not a technical restriction. Specifically, we are investigating the use of other ontology representation languages [11, 8].

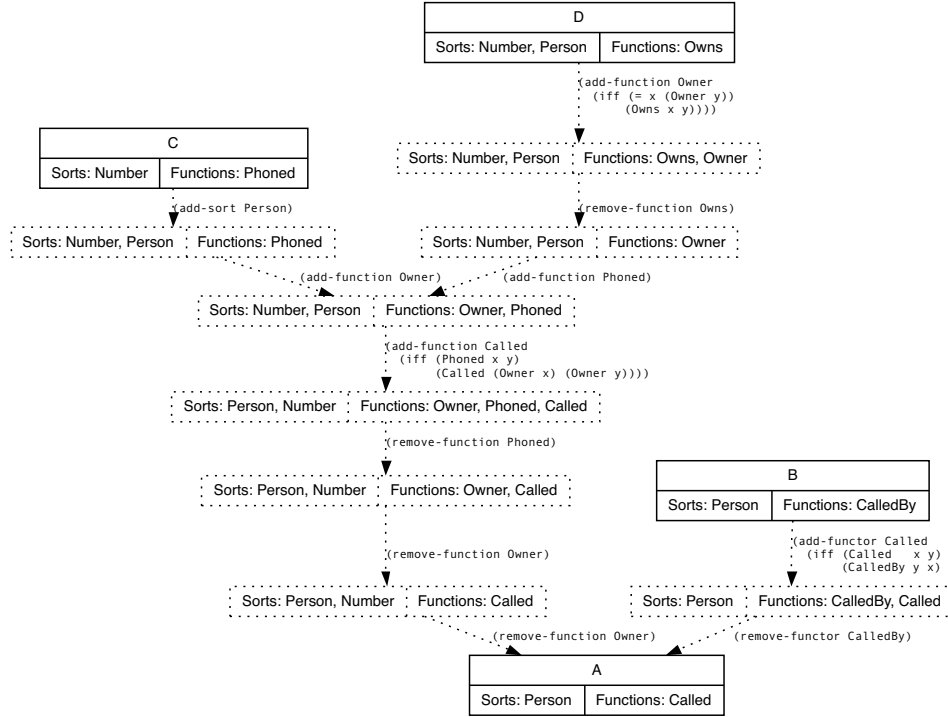


Figure 1: A sample translation graph enabling interoperability between four related ontologies.

2. *RemoveFunc*tor(OwnerOf)

3. *RemoveSort*(Firearms)

PBSI between the two described ontologies is made possible: Suppose that the first ontology has among the declarative information in its knowledgebase that Mohammed Al Harbi is the owner of an AKS-74U assault rifle, and that the knowledgebase of the second ontology contains no information about Mohammed Al Harbi except that he is a person. A query of whether or not Mohammed is armed, issued in the second ontology and making use of  $\sigma_1$ 's knowledgebase along with bridging axioms generated by traversing the path from  $\sigma_1$  to  $\sigma_2$ , would yield the correct answer and the associated, certifying proof.

It is important to note that PBSI provides a formal framework and corresponding implementation to break through the  $n^2$  barrier. In the case where translation between several ontologies is desired, translation graphs provide a means to surmount this  $n^2$  problem. This is achieved by use of an *intertheory* through which ontologies are interconnected thereby requiring only  $2n$  translation functions (see figure 2). Of course, an even bigger breakthrough would be secured if PBSI could be fully automated, and we turn no to that possibility.

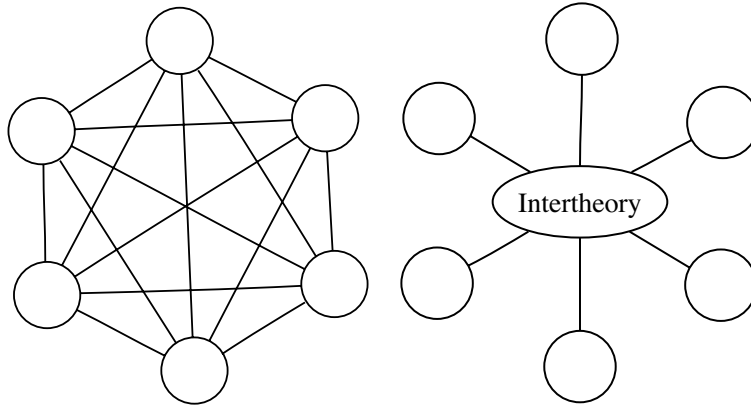


Figure 2: Interoperability between  $n$  ontologies (left) typically requires  $\binom{n}{2}$  connections but with an intertheory (right), interoperability is achieved using only  $2n$ .

## 4 Automation

In this section, we discuss ways to automate the process of creating and applying translation graphs. The procedure to extract appropriate bridging axioms from a translation graph has been accomplished, and systems whose ontologies are present as nodes in a translation graph can interoperate with other nodes in the graph. PBSI does not always yield *translation*; in some cases, bridging axioms can be converted to techniques for syntactic translation, but typically interoperability is achieved by a system issuing a query *expressed in its own syntax and semantics* and the search for an answer incorporates knowledge from related ontologies.

A detailed example of the above is presented in the interoperability experiment [2] between our own advanced reasoning system, Slate, and Oculus' geospatial and temporal visualization system, GeoTime. In the experiment, Slate and GeoTime collaborate to solve a portion of a case study used at the Joint Military Intelligence College. Additionally, the IKRIS Workshop [12] culminated in a demonstration of interoperability between three systems, Slate [1], Cycorp's Nöscape [14], and IBM and Stanford's KANI [5].<sup>2</sup>

This automation gets us half way there, but the holy grail of PBSI is to automate not only the intoperation between systems, but the generation of translation graphs as well. Translation graphs are of course implemented in code, so the challenge of fully automating PBSI<sup>+</sup> becomes the challenge of so-called *automatic programming* [13]. Because of the capability of the system we have designed for intelligence analysts (Slate), we are optimistic about being able to devise programs that generate the programs that implement translation graphs. Slate integrates deductive, inductive, and abductive reasoning. To the best of our knowledge, there has not been a single effort in automatic programming that synthesizes these three elements. The tradition of deductive program automation [10] is based *exclusively* on deduction; the tradition of machine learning (e.g., genetic programming [9]) is based *exclusively* on induction; while abduction has not even been

<sup>2</sup>Demonstrations of these experiments and other Slate-related content is made available online at <http://www.cogsci.rpi.edu/slate/Demos/>

explored in this field. And yet, typically, when humans approach a programming problem they employ all three of these. They use induction (in tandem with testing and checking) to formulate conjectures about the problem and their tentative solutions; they use deduction in order to reason about the consequences of their design decisions and about the correctness of their solutions; and they use abduction to explain the behavior of their algorithms. We look forward to reporting on our progress toward full automaticity at OIC 2007.

## 5 A Robust Example

In the presentation corresponding to this extended abstract at OIC 2007 itself, we will also describe a PBSI<sup>+</sup>-enabled interoperability example too robust to present within present space constraints. The example will be based on ongoing DTO-sponsored R&D, in which the aforementioned Oculus and Slate systems interoperate to enable analysts, working on a challenging case study, to issue hypotheses and recommendations that would not otherwise be attainable.

## References

- [1] BRINGSJORD, S., SHILLIDAY, A., AND TAYLOR, J. Slate, <http://www.cogsci.rpi.edu/slate/>, 2007.
- [2] CHAPPELL, A., BRINGSJORD, S., SHILLIDAY, A., TAYLOR, J., AND WRIGHT, W. Integration Experiment with GeoTime, Slate, and VIKRS. ARIVA Principal Investigator Meeting Handout, March 2007.
- [3] CHOI, N., SONG, I.-Y., AND HAN, H. A survey on ontology mapping. *SIGMOD Rec.* 35, 3 (2006), 34–41.
- [4] DOU, D., MCDERMOTT, D., AND QI, P. *Ontology Translation by Ontology Merging and Automated Reasoning*. Whitestein Series in Software Agent Technologies and Autonomic Computing. Birkhäuser Basel, 2005, pp. 73–94.
- [5] FIKES, R. E., FERRUCCI, D., AND THURMAN, D. A. Knowledge Associates for Novel Intelligence. In *Proceedings of the 2005 International Conference on Intelligence Analysis (IA 2005)* (McLean, VA, USA, May 2005).
- [6] GOGUEN, J. A. Data, Schema, Ontology and Logic Integration. *Logic Journal IGPL* 13, 6 (2005), 685–715.
- [7] HENDLER, J. Agents and the Semantic Web. *IEEE Intelligent Systems* (March/April 2001), 30–37.
- [8] KLYNE, G., AND CARROLL, J. J. Resource description framework (rdf): Concepts and abstract syntax, available at <http://www.w3.org/tr/rdf-concepts/>. Tech. rep., W3C, February 2004.

- [9] KOZA, J. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, 1992.
- [10] MANNA, Z., AND WALDINGER, R. Fundamentals of deductive program synthesis. In *Logic, Algebra, and Computation*, F. L. Bauer, Ed. Springer, Berlin, Heidelberg, 1991, pp. 41–107.
- [11] MCGUINNESS, D. L., AND VAN HARMELEN, F. OWL Web Ontology Language overview. Tech. rep., W3C, Available at <http://www.w3.org/TR/owl-features/>, 2004.
- [12] MITRE. IKRIS, workshop site: <http://nrrc.mitre.org/NRRC/ikris.htm>, 2007.
- [13] RICH, C., AND WATERS, R. C. Automatic programming: Myths and prospects. *Computer* 21, 8 (Aug. 1988), 40–51.
- [14] SIEGEL, N., SHEPARD, B., CABRAL, J., AND WITBROCK, M. Hypothesis Generation and Evidence Assembly for Intelligence Analysis: Cycorp’s Noöscape Application. In *Proceedings of the 2005 International Conference on Intelligence Analysis (IA 2005)* (McLean, VA, USA, May 2005).
- [15] SMITH, B. The basic tools of formal ontology. In *Formal Ontology in Information Systems* (1998), N. Guarino, Ed., IOS Press, pp. 19–28.
- [16] TAYLOR, J., SHILLIDAY, A., AND BRINGSJORD, S. Provability-based semantic interoperability via translation graphs. In *International Workshop on Ontologies and Information Systems for the Semantic Web (ONISW 2007)* (2007).





# **OIC-2007**

## **ONTOLOGY FOR THE INTELLIGENCE COMMUNITY**

### **TOWARDS EFFECTIVE EXPLOITATION AND INTEGRATION OF INTELLIGENCE RESOURCES**

#### **— ACKNOWLEDGEMENTS —**

#### **CONFERENCE DIRECTOR**

**Barry Smith**

National Center for Ontological Research - Buffalo  
University at Buffalo  
135 Park Hall, North Campus  
Buffalo, New York 14260  
phismith@buffalo.edu

---

#### **SCIENTIFIC COMMITTEE**

**Kathleen Stewart Hornsby (Chair)**

University of Iowa

**Bill Andersen**

Ontology Works, Inc.

**Selmer Bringsjord**

Rensselaer Polytechnic Institute

**Werner Ceusters**

New York State Center of Excellence  
in Bioinformatics & Life Sciences

**Randall Dipert**

University at Buffalo

**Terry Janssen**

Lockheed Martin Corp.

**Kathryn Laskey**

George Mason University

**Kevin Lynch**

Central Intelligence Agency

**Leo Obrst**

The MITRE Corporation

**Chris Welty**

IBM Watson Research Center

---

#### **ADDITIONAL REVIEWERS**

Richard MacMillan

Karla Massey

Dru (Donald) McCandless

Ken Samuel

Suzette Stoutenburg

The MITRE Corporation

A total of 33 papers were submitted for consideration for presentation at OIC-2007. With limited time slots available, the Scientific Committee was forced to make some very difficult decisions amongst several outstanding submissions. We wish to thank authors of all submitted papers for their time and interest in this event. An electronic version of the OIC-2007 proceedings is available on CEUR-WS (<http://CEUR-WS.org>).

---

***Sincere appreciation and gratitude is extended to an unnamed donor whose generous contribution to NCOR was used to fund portions of this event.***



