Horizontal Integration of Warfighter Intelligence Data

A Shared Semantic Resource for the Intelligence Community

Barry Smith
University at Buffalo, NY, USA

Tatiana Malyuta
Data Tactics Corp. VA, USA

William S. Mandrick
Data Tactics Corp. VA, USA

Chia Fu
Data Tactics Corp. VA, USA

Kesny Parent
Intelligence and Information Warfare Directorate (I²WD) CERDEC, MD, USA

Milan Patel
Intelligence and Information Warfare Directorate (I²WD) CERDEC, MD, USA

***Abstract* - We describe a strategy that is being used for the horizontal integration of warfighter intelligence data within the framework of the US Army's Distributed Common Ground System Standard Cloud (DSC) initiative. The strategy rests on the development of a set of ontologies that are being incrementally applied to bring about what we call the 'semantic enhancement' of data models used within each intelligence discipline. We show how the strategy can help to overcome familiar tendencies to stovepiping of intelligence data, and describe how it can be applied in an agile fashion to new data resources in ways that address immediate needs of intelligence analysts.**

***Index Terms*—semantic enhancement, ontology, joint doctrine, intelligence analytics, intelligence data retrieval.**

I. INTRODUCTION

The horizontal integration of warfighter intelligence data is described in Chairman of the Joint Chiefs of Staff Instruction J2 CJCSI 3340.02A [1] in the following way:

> Horizontally integrating warfighter intelligence data improves the consumers' production, analysis and dissemination capabilities. Horizontal Integration (HI) requires access (including discovery, search, retrieval, and display) to intelligence data among the warfighters and other producers and consumers via standardized services and architectures. These consumers include, but are not limited to, the combatant commands, Services, Defense agencies, and the Intelligence Community.

Horizontal integration is achieved when multiple heterogeneous data resources become aligned or harmonized in such a way that search and analysis procedures can be applied to their combined content as if they formed a single resource. We describe here a methodology that is designed to achieve such alignment in a flexible and incremental way. The methodology is applied to the source data at arm's length, in such a way that the data itself remains unaffected by the integration process.

Ironically, attempts to achieve horizontal integration have often served to consolidate the very problems of data stovepiping which they were designed to solve. Integration solution A is proposed; and works well for the data and purposes for which it was originally tailored; but it does not work at all when applied to new data, or to existing data that has to be used in new ways. Such failures arise for a variety of reasons, many of which have to do with the fact that integration systems are too closely tied to specific features of the (software/workflow) environments for which they have been developed. We propose a strategy for horizontal integration which seeks to avoid such problems by being completely independent of the processes by which the data store to which it is applied is populated and utilized. This strategy, which draws on standard features of what is now called 'semantic technology' [2], has been used successfully for over ten years to advance integration of the data made available to bioinformaticians, molecular biologists and clinical scientists in the wake of the successful realization of the Human Genome Project [3, 4]. The quantity and variety of such data – now spanning all species and species-interactions, at all life stages, at multiple granularity levels, and pertaining to thousands of different diseases – is at least comparable to the quantity and variety of the data which need to be addressed by intelligence analysts. As we describe in more detail in [5], however, today's dynamic environment of military operations (from Deterrence to Crisis Response to Major Combat Operations) is one in which ever new data sources are becoming salient to intelligence analysis, in ways which will require a new sort of agile support for retrieval, integration and enrichment of data. We will thus address in particular how our strategy can be rapidly reconfigured to allow its application to emerging data sources.

The strategy is one of a family of similar initiatives designed both to rectify the legacy effects of data stovepiping in the past and to counteract the problems caused by new stovepipes arising in the future. It is currently being applied within the DCGS-A Standard Cloud (DSC) initiative, which is part of the Distributed Common Ground System-Army [6], the principal Intelligence, Surveillance and Reconnaissance (ISR) enterprise for the analysis, processing and exploitation of all US Army intelligence data, and which is designed to be interoperable with other DCGS programs. The DSC Cloud is a military program of record in the realm of Big Data that is accumulating data from multiple diverse sources and with high rapidity of change. In [5, 7] we described how the proposed strategy is already helping to improve search results within the DSC Cloud in ways that bring benefits to intelligence analysts. In this communication, we present the underlying methodology describing also how it draws on resources developed in an incremental way that takes account of lessons learned in successive phases of application of the methodology to new kinds of data. Here we provide only general outlines. Further details and supplementary material are presented at [8].

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
OCT 2012		**00-00-2012 to 00-00-2012**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Horizontal Integration of Warfighter Intelligence Data:A Shared Semantic Resource for the Intelligence Community	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Intelligence and Information Warfare Directorate,Aberdeen Proving Ground,MD,21005**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Preprint, to be presented at SEMANTIC TECHNOLOGY FOR INTELLIGENCE, DEFENSE, AND SECURITY , (STIDS 2012), George Mason Univ, Fairfax, VA October 24-25, 2012

14. ABSTRACT
We describe a strategy that is being used for the horizontal integration of warfighter intelligence data within the framework of the US Army?s Distributed Common Ground System Standard Cloud (DSC) initiative. The strategy rests on the development of a set of ontologies that are being incrementally applied to bring about what we call the ?semantic enhancement? of data models used within each intelligence discipline. We show how the strategy can help to overcome familiar tendencies to stovepiping of intelligence data, and describe how it can be applied in an agile fashion to new data resources in ways that address immediate needs of intelligence analysts.

15. SUBJECT TERMS

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

II. Overcoming Semantic Stovepipes

Every data store is based on some data model which specifies how the data in the store is to be organized. Since communities that develop data stores do so always to serve some particular purpose, so each data model, too, is oriented around some specific purpose. Data models have been created in uncoordinated ways to address these different purposes, and they typically cannot easily be modified to serve additional purposes. Where there is a need to combine data from multiple existing systems, therefore, the tendency has been to invest what may be significant manual effort in building yet another data store, thereby contributing further to a seemingly never-ending process of data stovepipe proliferation.

To break out of this impasse, we believe, a successful strategy for horizontal integration must operate at a different level from the source data. It must be insulated from entanglements with specific data models and associated software applications, and it must be marked by a degree of persistence and of relative technological simplicity over against the changing source data to which it is applied.

The strategy we propose, which employs by now standard methods shared by many proponents of semantic technology [2], begins by focusing on the terms (labels, acronyms, codes) used as column headers in source data artifacts. The underlying idea is that it is very often the case that multiple distinct terms $\{t_1, \ldots, t_n\}$ are used in separate data sources with one and the same meaning. If, now, these terms are associated with some single 'preferred label' drawn from some standard set of such labels, then all the separate data items associated with the $\{t_1, \ldots t_n\}$ will become linked together through the corresponding preferred labels.

Such sets of preferred labels provide the starting point for the creation of what are called 'ontologies', which are created (1) by selecting a preliminary list of labels in collaboration with subject-matter experts (SMEs); (2) by organizing these labels into graph-theoretic hierarchies structured in terms of the *is_a* (or subtype) relation and adding new terms to ensure *is_a* completeness; (3) by associating logical definitions, lists of synonyms and other metadata with the nodes in the resultant graphs. One assumption widespread among semantic technologists is that ontology-based integration is best pursued by building large ontology repositories (for example as at [9]), in which, while use of languages such as RDF or OWL is standardized, the ontologies themselves are unconstrained. Our experience of efforts to achieve horizontal integration in the bioinformatics domain, however, gives us strong reason to believe that, in order to counteract the creation of new ('semantic') stovepipes, we must ensure that the separate ontologies are constructed in a collaborative process which ensures a high degree of integration among the ontologies themselves. To this end, our strategy imposes on ontology developers a common set of principles and rules and an associated common architecture and governance regime in order to ensure that the suite of purpose-built ontologies evolves in a consistent and non-redundant fashion.

III. Defining Features of the SE Approach

Associating terms used in source data with preferred labels in ontologies leads to what we call 'Semantic Enhancement' (SE) of the source data. The ontologies themselves we call 'SE ontologies', and the semantically enhanced source data together form what we call the 'Shared Semantic Resource' (SSR). To create this resource in a way that supports successful integration, our methodology must ensure realization of the following goals, which are common to many large-scale horizontal integration efforts:

- It must support an incremental process of ontology creation in which ontologies are constructed and maintained by multiple distributed groups, some of them associated with distinct agencies, working to a large degree independently.
- The content of each ontology must exist in both human-readable (natural language) and computable (logical) versions in order to allow the ontologies to be useful to multiple communities, not only of software developers and data managers, but also of intelligence analysts.
- Labels must be selected with the help of SMEs in the relevant domains. This is not because these labels are designed to be used by SMEs at the point where source data are collected; rather it is to ensure that the ontologies reflect the features of this domain in a way that coheres as closely as possible with the understanding of those with relevant expertise. Where necessary – for instance in cases where domains overlap – multiple synonyms are incorporated into the structure of the relevant ontologies to reflect usage of different communities of interest.
- Ontology development must be an arms-length process, with minimal disturbance to existing data and data models, and to existing data collection and management workflows and application software.
- Ontologies must be developed in an incremental process which approximates by degrees to a situation in which there is one single reference ontology for each domain of interest to the intelligence community.
- The ontologies must be capable of evolving in an agile fashion in response to new sorts of data and new analytical and warfighter needs.
- The ontologies must be linked together through logical definitions [10], and they must be maintained in such a way that they form a single, non-redundant and consistently evolving integrated network. The fact that all the ontologies in this network are being used simultaneously to create annotations of source data artifacts will in turn have the effect of virtually transforming the latter into an evolving single SSR, to

which computer-based retrieval and analysis tools can be applied.

The ontology development strategy we advocate thus differs radically from other approaches (such as are propounded in [11]), which allow contextualized inconsistency. For while of course source data in the intelligence domain will sometimes involve inconsistency – the data is derived, after all, from multiple, and variably reliable, sources –, to allow inconsistency among the ontologies used in annotations would, from our point of view, defeat the purposes of horizontal integration.

To achieve the goals set forth above, we require:

- A set of ontology development rules and principles, a shared governance and change management process, and a common architecture incorporating a common, domain-neutral, upper-level ontology.
- An ontology registry in which all ontology initiatives and emerging warfighter and analyst needs will be communicated to all collaborating ontology developers.
- A simple, repeatable process for ontology development, which will promote coordination of the work of distributed development teams, allow the incorporation of SMEs into the ontology development process, and provide a software-supported feedback channel through which users can easily communicate their needs, and report errors and gaps to those involved in ontology development.
- A process of intelligence data capture through 'annotation' [12] or 'tagging' of source data artifacts [7], whereby the preferred labels in the ontologies are associated incrementally with the terms embedded in source data models and terminology resources in such a way that the data in distinct data sources, where they pertain to a single topic, are represented in the SSR in a way that associates them with a single ontology term. Currently the annotation process is primarily manually driven, but it will in the future incorporate the use of Natural Language Processing (NLP) tools. Importantly, the process of annotation incrementally tests the ontologies against the data to which they must be applied, thereby helping to identify errors and gaps in the ontologies and thus serving as a vital ontology quality assurance mechanism [12].

IV. ONTOLOGICAL REALISM

The key idea underlying the SE methodology is that the successful application of ontologies to horizontal data integration requires a process for creating ontologies that is independent of specific data models and software implementations. This is achieved through the adoption of what is called 'ontological realism' [13], which rests on the idea that ontologies should be constructed as representations, not of data or of data models, but rather of the types of entities in reality to which the data relate.

The first step in the development of an ontology for a domain that has been identified as a target for intelligence analysis is thus *not* to examine what types of data we have about that domain. Rather, it is to establish in a data-neutral fashion the salient types of entities within the domain, and to select appropriate preferred labels for these types, drawing for guidance on the language used by SMEs with corresponding domain expertise. In addition, we rely on authoritative publications such as the capstone Joint Publication (JP) 1 of Joint Doctrine and the associated Dictionary (JP 1-02) [14, 15] (see Figure 1), applying adjustments where necessary to ensure logical consistency. The resultant preferred labels are organized into simple hierarchies of subtype and supertype, and each label is associated with a simple logical definition, along the lines illustrated (in a toy example) in Table 1.

vehicle =def: an object used for transporting people or goods
- **personnel carrier** =def. a vehicle that is used for transporting persons
- **tractor** =def: a vehicle that is used for towing
- **crane** =def: a vehicle that is used for lifting and moving heavy objects

vehicle platform=def. means of providing mobility to a vehicle
- **wheeled platform**=def. a vehicle platform that provides mobility through the use of wheels

Table 1. Fragments of asserted ontologies

V. REALIZATION OF THE STRATEGY

There is a tension, in attempts to create a framework for horizontal integration of large and rapidly changing bodies of data, which turns on the fact that (1) to secure integration the framework needs to be free from entanglements with specific data models; yet (2) to allow effective representation of data, the framework needs to remain as close as possible to those same data models.

This same tension arises also for the SE approach, where it is expressed in the fact that:

(1) The SSR needs to be created on the basis of persistent, logically well-structured ontologies designed to be reused in relation to multiple different bodies of data; yet:

(2) To ensure agile response to emerging warfighter needs, its ontologies must be created in ways that keep them as close as possible to the new data that is becoming available locally in each successive stage.

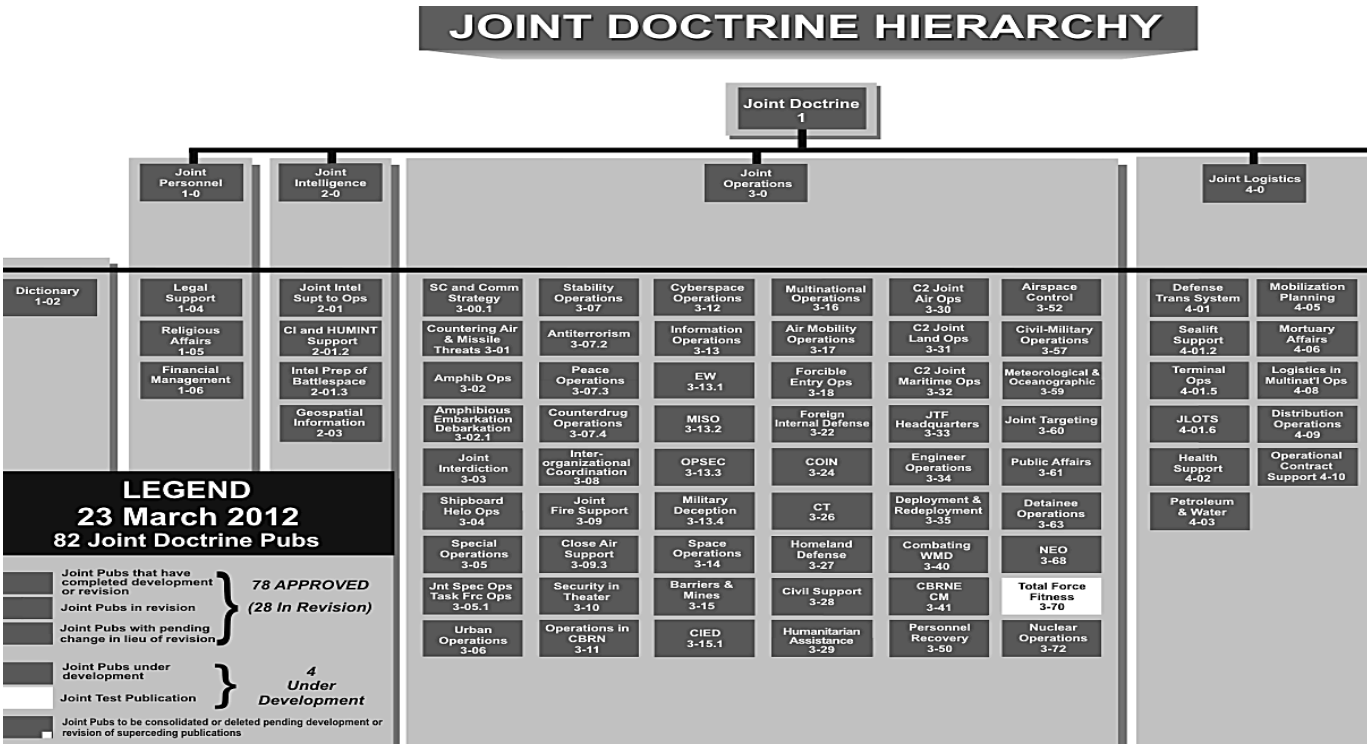

Figure 1 - Joint Doctrine Hierarchy

To resolve this tension, the SE strategy incorporates a distinction between two sorts of ontologies, called '*reference*' and '*application*' ontologies, respectively. By '*reference ontology*', we mean an ontology that captures generic content and is designed for aggressive reuse in multiple different types of context. Our assumption is that most reference ontologies will be created manually on the basis of explicit assertion of the taxonomical and other relations between their terms. By '*application ontology*', we mean an ontology that is tied to specific local applications. Each application ontology is created by using ontology merging software [16] to combine new, local content with generic content taken over from relevant reference ontologies [17,18], thereby providing rapid support for information retrieval in relation to particular bodies of intelligence data but in a way that streamlines the task of ensuring horizontal integration of this new data with the existing content of the SSR.

A. Principle of Single Inheritance

Our ontologies are 'inheritance' hierarchies in the sense that everything that holds (is true) of the entities falling under a given parent term holds also of all the entities falling under its *is_a* child terms at lower levels. Thus in Figure 2, for example, everything that holds of 'vehicle' holds also of 'tractor'. Each reference ontology is required to be created around an inheritance hierarchy of this sort that is constructed in accordance with what we call the *principle of asserted single inheritance*. This requires that for each reference ontology the *is_a* hierarchy is asserted, through explicit axioms (subclass axioms in the OWL language), rather than inferred by the reasoner. In addition it requires that this asserted *is_a* hierarchy is a monohierarchy (a hierarchy in which each term has at most one parent). This requirement is imposed for reasons of efficiency and consistency: it allows the total ontology structure to be managed more effectively and more uniformly across distributed development teams – for example by aiding positioning and surveyability of terms. It brings also computational performance benefits [23] and provides an easy route (described in Section V.E below) to the creation of the sorts of logical definitions we will need to support horizontal integration. The principle of asserted single inheritance comes at a price, however, in that it may require reformulation of content – for example deriving from multi-inheritance ontologies already developed by the intelligence community – that is needed to support the creation of the SSR. Again, our experience in the biomedical domain is that such reformulation, while requiring manual effort, is in almost all cases trivial, and that, where it is not trivial, the effort invested often brings benefits in terms of greater clarity as to the meanings and interrelationships of the new terms that need to be imported into the SE framework.

B. A Simple Case Study

Imagine, now, that there is a need for rapid creation of an application ontology incorporating preferred labels to describe artillery units available to some specific military unit called 'Delta Battery'. Such an ontology is enabled, first, by selecting from existing reference ontologies the terms needed to address the data in hand, for example of the sort used in Table 1. Second we define supplementary terms needed for our specific local case, as in Table 2.

Some of these terms may later be incorporated into corresponding asserted ontologies within the SE suite. For our present purposes, however, they can be understood as being simply combined together with the associated asserted ontology terms using ontology merging software, for example as developed by the Brinkley [17,19,17] and He [20,21] Groups. Because of the way the definitions are formulated, it is then possible to apply an automatic reasoner [22] to the result of merger to infer new relations, and thereby to create a new ontology hierarchy, as in Figure 2. Note that, in contrast to the reference ontologies from which it is derived, such an application ontology need not satisfy the principle of single inheritance. Note, too, that the definitions are exploited by the reasoner not only to generate the new inferred ontology, but also to test its consistency both internally and with the reference ontologies from which it is derived.

artillery weapon = def. device for projection of munitions beyond the effective range of personal weapons
artillery vehicle = def. vehicle designed for the transport of one or more artillery weapons
wheeled tractor = def. a tractor that has a wheeled platform
tracked tractor = def. a tractor that has a tracked platform
artillery tractor = def. an artillery vehicle that is a tractor
wheeled artillery tractor = def. an artillery tractor that has a wheeled platform
Delta Battery artillery vehicle=def. an artillery vehicle that is at the disposal of Unit Delta
Delta Battery artillery tractor=def. an artillery tractor that is at the disposal of Unit Delta

Table 2: Examples of supplementary terms and definitions

The strategy is designed to guarantee

(1) that salient reference ontology content is preserved in the new, inferred ontology in such a way that

(2) the latter can be used to semantically enhance newly added data very rapidly, and thereby

(3) bring about the horizontal integration of these data with all remaining contents of the SSR.

While ontology software has the capacity to support rapid ontology merger and consistency checking, we note that the inferred application ontology that is generated may on first pass fail to meet the local application needs. Thus, multiple iterations and investment of manual effort are needed.

Requiring that all inferred ontologies rest on reference ontology content serves not only to ensure consistency, but also to bring about what we can think of as the *normalization* [23] of the evolving ontology suite. (This is in loose analogy with the process of normalization of a vector space, where a *basis* of orthogonal unit vectors is chosen, in terms of which every vector in the whole space can be represented in a standard way.)

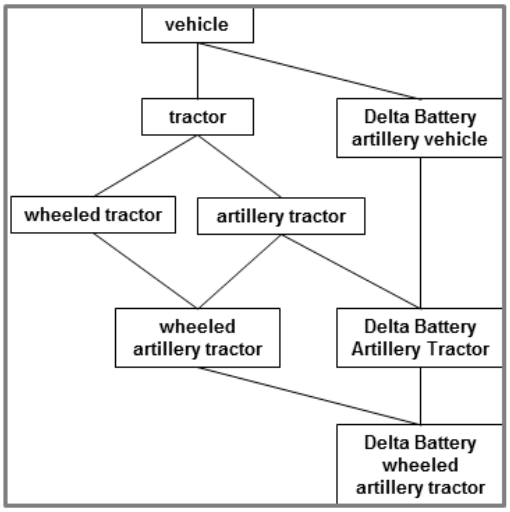

Figure 2. Inferred ontology of Delta Battery artillery vehicles. Child-parent links are inferred by the reasoner from the content of merged reference ontologies and from definitions of the supplementary terms. Note that some terms have multiple parents.

A suite of normalized ontologies is easier to maintain, because globally significant changes – those changes which potentially have implications across the entire suite of ontologies – can be made in just one place in the relevant reference ontology, thereby allowing consequent changes in the associated inferred ontologies to be propagated automatically. This makes ontology-based integration easier to manage and scale, because when single-inheritance modules serve to constrain allowable sorts of combinations, this makes it easier to avoid problems of combinatorial explosion.

C. Modularity of Ontologies Designed for Reuse

The reference ontologies within the SE suite are to be conceived as forming a set of plug-and-play ontology modules such as the Organization Ontology, Geospatial Feature Ontology, Human Physical Characteristics Ontology, Event Ontology, Improvised Explosive Device Component Ontology, and so on. These modules need to be created at different levels of generality, with the architecture of the higher level reference ontologies being preserved as we move down to lower levels.

Each module has its own coverage domain, and the coverage domains for the more specific modules (for example *artillery vehicle*, *military engineering vehicle*) are contained as parts within the coverage domains of the more general modules (for example *vehicle*, *equipment*). It is our intention that the full SE suite of ontologies will mimic the sort of hierarchical organization that we find in the Joint Doctrine Hierarchy [15], and our strategy for identifying and demarcating modules will wherever possible follow the demarcations of Joint Doctrine. The goal is to specify a set of levels of greater and lesser generality: for example *Intelligence*, *Operations*, *Logistics*, at one level; *Army Intelligence*, *Navy Intelligence*, *Airforce Intelligence*, at the next lower level; and so on. Ideally, the set of modules on

each level are non-redundant in the sense that (1) they deal with non-overlapping domains of entities and thus (2) do not contain any terms in common. In this way the more general content at higher levels is inherited by the lower levels and thus does not need to be recreated anew. As the history of doctrine writing shows, drawing such demarcations and ensuring consistency of term use in each sibling domain on any given level is by no means easy. Here, however, we will have the advantage that the ontology resource we are creating is not designed to serve as a terminology and doctrine set for use by multiple distinct groups of warfighters. Rather, it is designed for use behind the scenes for the specific purpose of data discovery and integration. Thus it is assumed that disciplinary specialists will continue to use their local terminologies (and taxonomies) at the point where source data is being collected, even while, thanks to the intermediation of ontology annotation, they are contributing to the common SSR. At the same time, community-specific terms will wherever possible be added to the SE ontology hierarchies as synonyms. This will contribute not only to the effectiveness of ontology review by SMEs but also to the applicability of NLP technology in support of automatic data annotation.

Our goal is to build the SE ontology hierarchy in such a way as to ensure non-redundancy by imposing the rule that, for each salient domain, one single reference ontology module is developed for use throughout the hierarchy. Creating non-redundant modules in this way is, we believe, indispensable if we are to counteract the tendency for separate groups of ontology developers to create new ontologies for each new purpose.

D. Benefits of Normalized Ontology Modules

The grounding in modular, hierarchically organized, non-redundant, asserted ontology modules brings a number of significant benefits, of a sort which are being realized already in the biomedical ontology research referred to above [3]. First, it creates an effective division of labor among those involved in developing, maintaining and using ontologies. In particular, it allows us to exploit the existing disciplinary division of knowledge and expertise among specialists in the domains and subdomains served by the intelligence community. To ensure population of the ontologies in a consistent fashion, we are training selected SMEs from relevant disciplines in ontology development and use; at the same time we are ensuring efficient feedback between those who are using ontologies in annotating data and those who are maintaining the ontologies over time in order to assure effective update, including correction of gaps and errors.

Second, it ensures that the suite of asserted ontologies is easily surveyable: developers and users of ontologies can easily discover where the preferred label equivalents of given terms are to be found in the ontology hierarchy; they can also easily determine where new terms, or new branches, should be inserted into the SE suite. Thus, where familiar problems arise when mergers are attempted of independently developed ontologies and terminology content, the incremental approach adopted here implies that mergers will be applied almost exclusively only (1) to the content of reference ontologies developed according to a common methodology and reviewed at every stage for mutual consistency and (2) to application ontology content developed by downward population from the evolving ontology suite.

E. Creating Definitions

The principle of single inheritance allows application of a simple rule for formulating definitions of ontology terms, whereby all definitions are required to have the form:

an *S* = Def. a *G* which *D*s

where ‘S’ (for: species) is the term to be defined, ‘G’ (for: genus) is the immediate parent term of ‘S’ in the relevant SE asserted ontology, and ‘D’ (for: differentia) is the species-criterion, which specifies what it is about certain G’s which makes them S’s. (Note that this rule can be applied consistently only in a context where every term to be defined has exactly one asserted parent.)

As more specific terms are defined through the addition of more detailed differentia, their definitions encapsulate the taxonomic information relating the corresponding type within the SE ontology to the sequence of higher-level terms by which it is connected to the corresponding ontology root. The task of formulating definitions thereby serves as a quality control check on the correctness of the constituent hierarchies, just as awareness of the hierarchy assists in the formulation of coherent definitions.

A further requirement is that the definitions themselves use (wherever possible) preferred labels which are taken over from other ontologies within the SE suite. Where appropriate terms are missing, the SE registry serves as a feedback channel through which the corresponding need can be transmitted to those tasked with ontology maintenance. The purpose of this requirement is to bring it about that the SE ontologies themselves will become incrementally linked together via logical relations in the way needed to ensure the horizontal integration of the data in the SSR that have been annotated with their terms. And as more logical definitions are added to the SE suite, the more its separate modules begin to act like a single, integrated network. All of this brings further benefits, including:

- Lessons learned in experience developing and using one module can be easily propagated throughout the entire system.
- The value of training in ontology development in any given domain module is increased, since the results of such training can easily be re-applied in relation to other modules.
- The incrementally expanding stock of available reference ontology terms will help to make it progressively easier to create in an agile fashion new application ontologies for emerging domains.

- The expanding set of logical definitions cross-linking the ontologies in the SE suite will mean that the use of ontology reasoners [22] for quality assurance of both asserted and inferred ontologies will become progressively more effective. These same reasoners will then be able to be used to check the consistency of the resultant annotations; and when inconsistencies are detected, these can be flagged as being of potential significance to the intelligence analyst.

VI. From Data to Decisions: An Example

Suppose, for example, that analysts are faced with a large body of new data pertaining to activities of organizations involved in the financing of terrorism through drug trafficking. The data is presented to them in multiple different formats, with multiple different types of labels (acronyms, free text descriptions, alphanumeric identifiers) for the types of organizations and activities involved.

To create a semantically enhanced and integrated version of these data for purposes of indexing and retrieval, analysts and ontology developers can use as their starting point the Organization Ontology which has already been populated with many of the general terms they will need across the entire domain of organizations, both military and non-military, formal and informal, family- or tribe- or religion-based, and so on. It will also contain the terms they need to define different kinds of member roles, organizational units and sub-units, chains of authority, and so on.

Adherence to the SE principles ensures that the Organization Ontology has been developed in such a way as to be interoperable, for example, with the Financial Event and Drug Trafficking Ontologies. Portions of each of these modules can thus be selected for merger in the creation of a new, inferred ontology, which can rapidly be applied to annotation of the new drug-financed terrorism data, which thereby becomes transformed from a mere collection of separate data sources into a single searchable store horizontally integrated within the SSR.

VII. Upper-, Mid-and Lowest-Level Ontologies

The SE suite of ontologies is designed to serve *horizontal* integration. But, it depends also on what we can now recognize as a *vertical* integration of asserted ontologies through the imposition of a hierarchy of ontology levels. In general, the SE methodology requires that all asserted ontologies are created via downward population from a common top-level ontology, which embodies the shared architecture for the entire suite of asserted ontologies – an architecture that is automatically inherited by all ontologies at lower levels.

Here, the *level* of an ontology is determined by the level of generality of the types in reality which its nodes represent. The Upper Level Ontology (ULO) in the SE hierarchy must be maximally general – it must provide a high-level domain-neutral representation of distinctions between objects and events, objects and attributes, roles, locations, and so forth. For this purpose we select the Basic Formal Ontology 2.0 (BFO), which has been thoroughly tested in multiple application areas [8,24]. Its role is to provide a framework that can serve as a starting point for downward population in order to ensure consistent ontology development at lower levels. Since almost all SE ontology development is at the lower levels within the hierarchy, BFO itself will in most cases be invisible to the user.

The Mid-Level Ontologies (MLOs) introduce successively less general and more detailed representations of types which arise in successively narrower domains until we reach the Lowest Level Ontologies (LLOs). These LLOs are maximally specific representation of the entities in a particular one-dimensional domain, as illustrated in Table 3.

Some MLOs are created by adding together LLO component modules, for example, the Person MLO may be created by conjoining person-relevant ontology components from Table 3 such as: Person Name, Person Date, Hair Color, Gender, and so on. More complex MLOs will involve the use of reasoners to generate ontologies incorporating inferred labels such as 'Male Adult', 'Female Infant', and so on, along the lines sketched in Section V.B above.

Person Name (with types such as: FirstName, LastName, …)
Hair Color (with types such as Grey, Blonde, …)
Military Role (with types such as: Soldier, Officer, …)
Blood Type (with types: O, A, …)
Eye Color (with types: Blue, Grey, …)
Gender (with types: Male, Female, …)
Age Group (with types: Infant, Teenager, Adult, …)
Person Date (with types: BirthDate, DeathDate, …)
Education History (with types: HighSchoolGraduation, …)
Education Date (with types: DateOfGraduation, …)
Criminal History (with types: FirstArrest, FirstProsecution, …)
Citizenship (based on ISO 3166 Country Codes)

Table 3. Examples of Lowest Level Ontologies (LLOs)

Figure 3 illustrates the rough architecture of the resultant suite of SE ontologies on different levels, drawing on the top-level architecture of Basic Formal Ontology.

VIII. Conclusion

In any contemporary operational environment, decision makers at all levels, from combatant commanders to tactical-level team leaders, need timely information pertaining to issues ranging from insurgent activity to outbreaks of malaria and from key-leader engagements to local elections. This requires the exploitation by analysts of a changing set of highly disparate databases and other sources of information, whose horizontal integration will greatly facilitate this data to decision cycle.

The SE strategy is designed to create the resources needed to support such integration incrementally, with thorough testing at each successive stage, and one of our current pilot projects is designed to identify the problems which arise when the SE methodology is applied to support

collaboration across distinct intelligence agencies, including exploring how independently developed legacy ontologies can be incorporated into the framework.

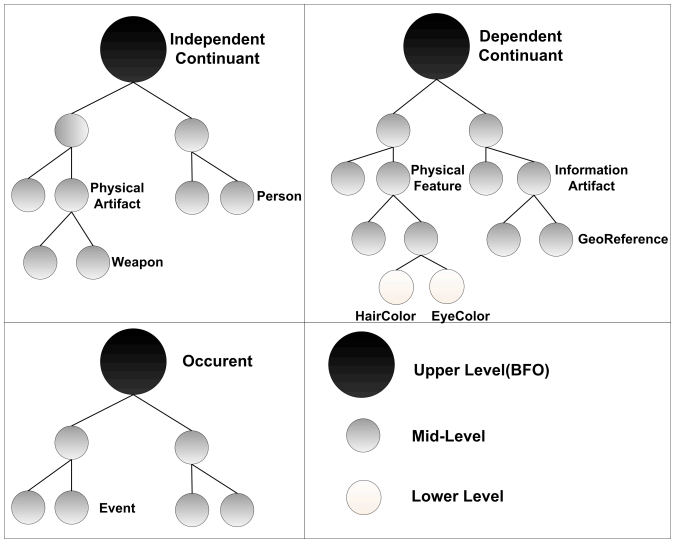

Figure 3. Organization of asserted ontologies

Our work on using SE ontologies for purposes of annotation has been executed thus far both manually and with NLP support. The results of this work have been found useful to indexing and retrieval of large bodies of data in the DSC Cloud store. In our next phase we will test its capacity to support rapid creation of application ontologies to address emerging analyst needs. In a subsequent, and more ambitious phase, we plan to explore the degree to which the idea of semantic enhancement can be truly transformative in the sense that it will influence the way in which source data are collected and stored. We believe that such an influence would bring a series of positive consequences flowing from the fact that the asserted ontologies will be focused automatically upon (i.e. represent) the same entities in the battlespace that the operators, analysts, and war-planners are concerned with, and they would treat these entities in the same intuitively organized way. Thus while at this stage all SE ontologies are free of entanglements with specific source data models, our vision for the future is that the success of the approach will provide ever stronger incentives for the use of SE ontologies already in the field. These incentives will exist, because using such ontologies at the point of data collection will guarantee efficient horizontal integration with the contents of the SSR, thereby giving rise to a network effect whereby not only the immediate utility of the collected data will be increased, but so also will the value of all existing data stored within the SSR.

REFERENCES

[1] Chairman of the Joint Chiefs of Staff Instruction. J2 CJCSI 3340.02A.

[2] P. Hitzler, M. Krötzsch and S. Rudolph, *Foundations of Semantic Web Technologies*, Chapman & Hall, 2009.

[3] Barry Smith, et al., "The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration", *Nature Biotechnology*, 25 (11), November 2007, 1251–1255.

[4] Fahim T. Imam, et al., "Development and use of Ontologies Inside the Neuroscience Information Framework: A Practical Approach", *Frontiers in Genetics*, 2012; 3: 111.

[5] Barry Smith, et al., "Ontology for the Intelligence Analyst", *Crosstalk: The Journal of Defense Software Engineering* (forthcoming).

[6] Distributed Common Ground System - Army (DCGS-A) What is it? *Pentagon Army Posture Statement*, 27 December 2011.

[7] David Salmen, et al., "Integration of Intelligence Data through Semantic Enhancement", *Proceedings of the Conference on Semantic Technology in Intelligence, Defense and Security* (STIDS), George Mason University, Fairfax, VA, November 16-17, 2011, CEUR, Vol. 808, 6–13

[8] Supplementary material on Semantic Enhancement: http://ncorwiki.buffalo.edu/index.php/Semantic_Enhancement

[9] http://ontolog.cim3.net/cgi-bin/wiki.pl?OpenOntologyRepository.

[10] Chris J. Mungall et al., "Cross-product extensions of the Gene Ontology", *Journal of Biomedical Informatics* 44 (2007), 80–86.

[11] Douglas B. Lenat, "CYC: a large-scale investment in knowledge infrastructure", *Communications of the ACM*, 38 (11), 1995 33-38.

[12] David P. Hill, et al., "Gene Ontology Annotations: What they mean and where they come from*", BMC Bioinformatics*, 2008; 9(Suppl 5): S2.

[13] Barry Smith and Werner Ceusters, "Ontological Realism as a Methodology for Coordinated Evolution of Scientific Ontologies", *Applied Ontology*, 5 (2010), 139–188.

[14] Joint Publication 1, Doctrine for the Armed Forces of the United States, Chairman of the Joint Chiefs of Staff. Washington, DC. 20 March 2009.

[15] Joint Electronic Library: The Joint Publications.

[16] Z. Xiang, et al., "OntoFox: Web-Based Support for Ontology Reuse", *BMC Research Notes*. 2010, 3:175.

[17] Marianne Shaw, et al., "Generating Application Ontologies from Reference Ontologies", *Proceedings, American Medical Informatics Association Fall Symposium*, 2008, 672-676.

[18] James Malone and Helen Parkinson, "Reference and Application Ontologies."

[19] James F. Brinkley et al., "Project: Ontology Views."

[20] http://www.hegroup.org/ontoden/.

[21] J. Hur, et al., "Ontology-based Brucella vaccine literature indexing and systematic analysis of gene-vaccine association network", *BMC Immunology* 2011, 12:49

[22] OWL 2 Reasoners, http://www.w3.org/2007/OWL/wiki/Implementations.

[23] Rector, A. L. "Modularisation of Domain Ontologies Implemented in Description Logics and Related Formalisms including OWL". *Proceedings of the 2nd International Conference on Knowledge Capture*, ACM, 2003, 121–128.

[24] Pierre Grenon and Barry Smith, "SNAP and SPAN: Towards Dynamic Spatial Ontology", *Spatial Cognition and Computation*, 4: 1 (March 2004), 69–103.