# DISTRIBUTED REPOSITORIES OF REASONABLE ONTOLOGIES FOR SNE COMPONENT RETRIEVAL AND REUSE


| Dr. Levent Yilmaz | Julio De La Cruz | Thai Nguyen | Todd Kohler |
|---|---|---|---|
| Trident Systems Incorporated | U.S. Army STRICOM | U.S. Army STRICOM | U.S. Army STRICOM |
| Fairfax, VA | Orlando, FL | Orlando, FL | Orlando, FL |


## ABSTRACT


Driven by fiscal constraints, increasing pressure exists to employ simulation for driving exploration into new and more effective methods for next generation modeling activities. The shorter timescales required as well as the need to rapidly incorporate operational elements demanded by these applications bring new challenges. In particular, cost reduction in Synthetic Natural Environment (SNE) development using parameterized scenario generation with compositional modeling is one of the core problems that need to be addressed. In particular, component-based development of complex SNE components requires domain-specific repository organization with intelligent component retrieval mechanisms along with reasoning methods for conflict identification before component assembly. Furthermore, integrating disparate SNE components needs sophisticated metadata interoperability analysis, diagnosis, and merge tools. In this paper, a brief overview of an ontology-oriented and web-based SNE retrieval and fusion framework is discussed. The proposed approach aims to support component-based SNE development lifecycle. Specifically, DARPA Agent Markup Language (DAML), the ontology language of Semantic Web, is proposed to uniformly specify and facilitate reasoning about the service profiles of environmental models as well as SNE data components. The framework builds on DAML and constitutes: (1) an ontology engineering subsystem, (2) an agent-based distributed metadata (ontology) broker infrastructure, built on Trident's Interchange[SE] - Object Oriented Data Repository, for SNE publishing, discovery, and retrieval, (3) diagnosis methods for SNE component certification before publishing, and (4) compositional consistency analysis and verification methods for integrated SNE component assemblies.


## ABOUT THE AUTHORS

**DR. LEVENT YILMAZ** is a senior research engineer at the Simulation and Software Division of Trident Systems Incorporated and an adjunct assistant professor at the Graduate School of Information and Telecommunication Studies of University of Maryland (UC). He holds M.S. and Ph.D. degrees in Computer Science from Virginia Polytechnic Institute and State University. His current research focuses on bringing compositional verification and reuse technologies along with formal models of behavioral consistency analysis to parameterized model and SNE retrieval, composition, and interoperability. Dr. Yilmaz is a member of ACM, IEEE, and SCS.

**JULIO DE LA CRUZ** is a project engineer at U.S. Army Simulation, Training, and Instrumentation Command (STRICOM) in Orlando, Florida. He has fourteen years of experience working within the DoD/Army acquisition process. He has been a lead project engineer for Major Defense Acquisition Programs (ACAT I, II, III) at various points of his civil servant career since 1988.

**THAI H. NGUYEN** is a principal investigator at U.S. Army Simulation, Training, and Instrumentation Command (STRICOM) in Orlando, Florida. He has fourteen years of experience working within NASA and the DoD. He has been a project engineer for Major Defense Acquisition Programs, a senior engineer for Boeing and software engineer for the Navy at various points of his engineering career since 1988.

**DOUGLAS T. KOHLER** is a Principal Investigator at U.S. Army Simulation, Training, and Instrumentation Command (STRICOM) in Orlando, Florida. He has fifteen years of experience with Major Defense Acquisition Programs for the Air Force, Navy and the Army. Currently Mr. Kohler is a principal investigator for the Synthetic Natural Environments (SNE) program in the ARMY Synthetic Environment & Technology Management Division.
.

# DISTRIBUTED REPOSITORIES OF REASONABLE ONTOLOGIES FOR SNE COMPONENT RETRIEVAL AND REUSE


| Dr. Levent Yilmaz | Julio De La Cruz | Thai Nguyen | Todd Kohler |
|---|---|---|---|
| Trident Systems, Incorporated | U.S. Army STRICOM | U.S. Army STRICOM | U.S. Army STRICOM |
| Fairfax, VA | Orlando, FL | Orlando, FL | Orlando, FL |


## Introduction

The need and demands for military simulation are increasing. Driven by fiscal constraints, increasing pressure exists to employ simulation for driving exploration into new, more effective methods for modeling combat activities. One of the primary reasons is increasing resource limitations such as the availability of actual equipment for training purposes, costs associated with live training environments, and safety as well as security reasons that could hinder mission training in real operating scenarios. DOD is currently investing and focusing significant amount of attention in the areas of *command control* system and *model* integration for mission planning, execution monitoring, and embedded training.

Over the years large numbers of SNE and model databases have been developed for analysis, distributed mission training, and many other simulation applications by the M&S communities. These databases constitute both terrain data and environmental models for SNE that are composed of atmospheric, space, oceanographic, and geological components. Recently, DMSO initiated the Integrated Natural Environment Authoritative Representation Process (INEARP) that aims to create a SNE that is physically consistent, both within and among the atmosphere, ocean, space, and terrain domains, using such COTS SNE components. Similarly, a recent initiative by STRICOM and SAIC has established a testbed for a repository of basic SNE objects to facilitate reuse and component-based development of SNE. The SNE Virtual Data Repository (SVDR) initiative provides a framework where ideas in retrieval and exchange of SNE products can be analyzed.

Note, however, integrated and compositional development of complex SNE requires formal means to retrieve potentially reusable components along with mechanisms for identification of conflicts among them. Furthermore, merging and integrating these disparate components to form a coherent and consistent SNE need sophisticated schema interoperability analysis, diagnosis, and merge tools. Existing retrieval methods and testbeds facilitate component location using limited component interface/feature definition and vocabularies without imposing logical domain constraints on the features of components to be retrieved. Locating SNE information on heterogeneous repositories, determining the usability of retrieved SNE components, and diagnosing and merging these components require a more sophisticated semantic analysis framework beyond the capabilities offered by the existing testbed. The proposed approach and viewpoints have potential to build on top the existing testbed by extending it with semantic ontology processing with formal analysis mechanisms and utilities.

Basically, the goal of this paper is to discuss and argue about how an ontology-based approach with sound formal semantics can support a component-based approach for SNE development to address the above problem.

## Compositional and Parameterized Construction of Integrated SNEs

Consider the following scenarios that help motivate the need for applications envisaged in this proposed research and development effort.

**Scenario 1:** A SNE designer, given the task of developing the terrain data or environmental model federation for a particular scenario, instantiates a target domain schema and then using a metadata search tool automatically identifies a set of SNE ontologies from a repository that syntactically and semantically match the required *concepts* and *constraints*. The engineer then retrieves the terrain data or models that are associated with the selected schemas.

**Scenario 2:** Before integrating models to a given federation or compiling the data sets to composite master visual database, the SNE designer uses meta-data exploration tools (*ontology interpretation and* *translation tools*) to relate the schema concepts to find inter-schema inconsistencies and correlate the data resources by (1) using their schemas and user-defined referential integrity constraints and (2) applying the domain-specific ontology-based consistency analysis heuristics to identify thematic layer conflicts. Finally, the individual data stores are integrated and translated, according to the mappings derived by the *ontology translation tool*, into the SEDRIS transmittal format.

**Scenario 3:** Before submitting the new composite SNE component to the repository for reuse, the SNE designer verifies the logical consistency of the composite metadata (schema) and demonstrates the conformance of the actual SNE component with respect to its published metadata. Then the designer validates the proper usage and conformance to envisioned application scenario by instantiating a set of ground truth SNE environmental models that effect, impact, and manipulate the internal dynamics of the terrain. The SNE designer generates a federation object model and retrieves a set of ground truth models from HLA model database, checks the consistency of the federation interactions and data models and instantiates the models to perform specific interaction and data manipulation tests on SNE terrain data.

## Rapid Compositional Building of an Integrated, Common, Consistent Simulated Environment

A simplified point of view regarding the life-cycle of component-based SNE development is depicted in Figure 1. The first phase of this lifecycle involves *component qualification/acquisition*, where components that satisfy the customer requirements are retrieved from libraries and external vendors. Next, as depicted in figure 1, the *component assembly* stage facilitates the integration of these components into a cohesive new SNE federation. Continuous *component analysis* determines the fitness and compatibility of the components and determines if reused components fulfill expected constraints. Modifications are made until the component is judged acceptable.

A set of concrete problems that need to be addressed for SNE retrieval and fusion are as follows:

- *Issues that pertain to component-based SNE development process:* What are the distinguishing characteristics of a component-based SNE development process? With what features do we need to augment STRICOM SNE Virtual Data Repository (SVDR) vision to support this process?
- *Issues that pertain to SNE component discovery and retrieval:* How can we augment SEDRIS DRM to define SNE component semantics along with logical feature constraints to facilitate constraint-based retrieval? What are the components of an intelligent ontology broker architecture that facilitates discovery, retrieval, and distributed registry of components?
- *Issues that pertain to SNE composition diagnosis for intra/inter domain inconsistencies:* How do we measure the degree with which the retrieved SNE matches user requirements? What domain specific properties are relevant to characterize inter/intra-domain consistency? Which modeling perspective and semantic domain is appropriate to formalize these properties and what are the proper analytical methods of this domain to compute if conditions that satisfy these properties exist? Finally, how do we determine if an actual SNE component conforms to its metadata (ontology) specification before published in the repository?

## Why is this Problem Important?

The major programs that will be supported by effective SNE representation are advance concept development and defense planning, R&D, acquisition, and military doctrine development. The GIS and C4I fields, for example, are also increasingly requiring the use of 2D and 3D data for visualization and evaluation in support of decision-making tasks.

Besides, there are commercial applications of the proposed technology in the knowledge management domain such as web commerce, web portals, and retrieval and assembly of multimedia collections.

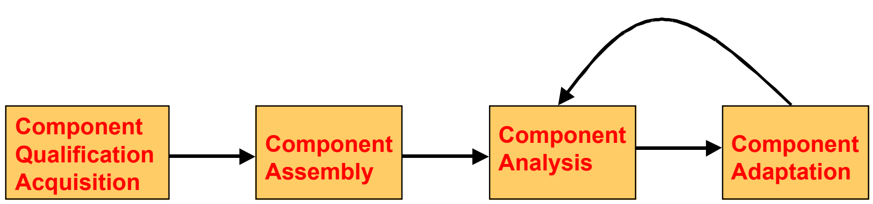


**Figure 1: Simplified Compositional Development Process**

## What is Required?

Component-based development of complex SNE requires domain-specific repository organization with intelligent component retrieval mechanisms along with reasoning methods for conflict identification before component assembly. Furthermore, merging disparate SNE components needs sophisticated schema interoperability analysis, diagnosis, and merge tools.

## Relevance of the Problem to the INEARP and the SVDR Initiatives

Recently SAIC and STRICOM initiated a new project, called SVDR, to promote establishing SNE repositories. The goal is to facilitate retrieval and reuse of SNE objects for rapid development of synthetic environment. Under funding from STRICOM, SAIC is establishing a test bed for the development of a repository, which will provide a framework and an environment where ideas for classification, retrieval, and exchange of SNE products can be developed and tested. Among the ideas being investigated are issues such as metadata content standards, access mechanisms to limited distribution databases and models, user requirements (both database content and search engines), and linkages to SEDRIS [Schaefer *et al.* 2002].

Meanwhile, Integrated Natural Environment Authoritative Representation Process (INEARP) project that is initiated by DMSO aims to create a natural environment representation that is physically consistent, both within and among the atmosphere, ocean, space, and terrain domains, for delivery to a customer application / system that meets the needs of its components. High-level descriptions of the envisioned INEARP technologies are follows: *Library Services* through the Master Environmental Library (MEL) facilitates discovery, access, subscription, and delivery of environmental information, products, and data. *Scenario Composition and Scenario Generation* through the Environmental Scenario Generator (ESG) provides the customer interface for scenario composition and scenario generation. The Interface will translate the customers' environmental requirements (e.g., rocky, rainy, high waves, solar disturbances) into environmental parameters (e.g., temperature, wind, sea state, etc., to include determination of the appropriate scale of the needed parameters to determine the Authoritative Representation) [DMSO 2000]

## The Relevance of the Problem to the INEARP and SVDR Challenges

The retrieval of relevant SNE components in the context of continuously changing dynamic information resources needs flexible, autonomous, intelligent publishing and retrieval mechanisms. Furthermore, the reusability of retrieved components in the application domain of interest requires identification of conflicts of the reused components with respect to the context they are embedded. Composition of retrieved SNE components and overlaying them on the terrain skin requires sophisticated schema-based consistency analysis methods that will augment the existing testbed. Our position is that performing consistency analysis at higher levels of SNE abstraction would facilitate avoiding costly data correlation until feasible and reusable components are identified with fast abstract correlation mechanisms based on ontology-based specification matching algorithms.

## Approach: Distributed Repositories of Reusable, and Reason-able Ontologies for Web-Based SNE Fusion

In this section we outline the envisioned approach in the context of an ongoing project. Section 4.1 discusses the component qualification requirements of a component-based development process and how an ontology-based approach could support these requirements. Section 4.2 introduces the proposed ontology infrastructure along with discussions about the underlying rational. Section 4.3 focuses on SNE repository organization and component retrieval, while section 4.4 discusses SNE ontology composition and diagnosis.

### Supporting SNE Component Reuse with Component Repositories

Component-based modeling is inherently a compositional process in which retrieved reusable components are adapted and assembled to form new composite components that provide the required services imposed by the domain-specific requirements. Component-based SNE modeling process constitutes component qualification, adaptation, assembly, and evolution stages. Hence, structured SNE component repositories are central to component-based development as they contribute to effective component qualification and retrieval to facilitate reuse.

Our investigation into component repository organization and structuring led to three main approaches originated from the information and multimedia retrieval research: (1) hierarchical classification, (2) aspect-oriented classification, and (3) indexing.

In *hierarchical classification* components are enumerated in categories that are usually structured in subcategories. The issues involved in using an enumerated classification include its inherent rigidity and problems with understanding large hierarchies. There is a tradeoff between the depth of a hierarchy and the number of category members. Some domains lend themselves to many small classes. The effect is that users unfamiliar with its organization and configuration will become lost in the chaos of class organizations. Other domains will have only few numbers of categories, but must inescapably contain many members. In this case, selection of a class is only a first step in the retrieval process, as the user must then explore a large number of category members for relevant information. Furthermore, the hierarchy gives only one aspect of the repository. Changes to that aspect may cause implications leading to extensive reconfiguration of class structures that can have consequences for the entire contents of the repository. Besides, hierarchical classification requires users to be familiar with the domain to effectively retrieve relevant components. Part of the problem is that most retrieval systems assume that an information space can be adequately represented with a single classification. But no one classification is correct under all circumstances, and it is impossible in principle to identify all possible relevant features of a large information space.

In *aspect-oriented classification* concepts are categorized into fixed number of mutually exclusive aspects. Users seek for components by expressing terms for each aspect. This method resembles and is a variation of relational model in which concepts are categorized according to attribute-value structures as in frame-based retrieval techniques of artificial intelligence. Since aspects can be redesigned without impact on other facets, aspect-oriented classifications are more flexible than enumerated classifications. While aspects make it easy to fuse and combine terms to represent components, it becomes hard for users to find the right combination of terms that accurately describe the information need, especially in complex information spaces. The method also requires that users know how the library and terms are structured and have an understanding of the significance of each facet and the terms that are used in the aspect.

*Indexing methods* use the keywords from component metadata store for indexing. Component metadata is applied to a "stop list" to remove frequently occurring words such as "and" and "the." The remaining attributes are used as an index to the repository. Users specify a query using keywords that are applied to the indices to find matching components. No classification effort is required, although human indexers are sometimes used to augment automatically extracted index terms. Matching criteria can range from Boolean match to more sophisticated methods, such as the vector model, that use statistical measures to rank retrieved information.

Later in this section, we propose a preliminary agent-based solution to repository organization that synergistically combines the above schemes with a service advertisement and discovery protocol to increase the effectiveness, scalability, and autonomy of SNE component repository management and retrieval. Furthermore, the proposed approach extends the power of classification methods by formal sound reasoning capabilities built on semantic metadata facilities, which are discussed in the next section. Hence, our position is that given a repository structure, the next challenge is the selection of a metadata that abstracts components to depict and advertise service profiles for retrieval purposes. Emergent standards such as XML and UML facilitate capturing domain structures and profiles of components; however, the lack of semantics in these standards enables only syntactic keyword-based retrieval of components. Besides, the lack of semantics rules out formal reasoning regarding constraint-based consistency analysis as well as component retrieval. In the next section we discuss the role of an ontology infrastructure in capturing the domain concept structure as well as the associated semantics for concept invariants, concept properties, and interrelationships among concepts.

## An Ontology-Based Infrastructure for SNE Component Abstraction

When SNE applications are developed compositionally using disparate SNE components, conflicts occur due to unanticipated implications of reused components in the new context. This happens because semantics of SNE behavioral models as well as data are rarely explicit. Hence, service requesters cannot determine if their assumptions are at least partially matched by reused components. Our position is that making these implicit assumptions explicit and accessible, as metadata would facilitate the conversion of semantic problems into syntactic problems, thus, would make such assumptions amenable to automated reasoning for conflict detection.

**Semantic Metadata (Ontology) Requirements for SNE:** A semantics-based ontology provides finite, but extensible vocabulary, unambiguous interpretation of classes and term relationships, strict hierarchical subclass relationships between classes. Furthermore, property and value restriction specification on per-class basis, specification of disjoint classes, distinguished relationships such as inverse, part-whole, as well as capability to specify arbitrary logical relationships between terms and concepts is a prerequisite for semantic ontology. Note however, most metadata representation models (i.e., UML in SEDRIS, XML) are syntactic and do not have a formal semantics. On the other hand, a recent DARPA initiative, called Semantic Web, resulted in a new ontology language with well-defined semantics. DAML+OIL is a semantic markup language for Web-based resources. It builds on earlier W3C standards such as RDF and RDF Schema, and extends these languages with richer modeling primitives. DAML+OIL provides modeling primitives commonly found in AI frame-based languages [Horrocks *et al*. 2001].

The language has a well-defined semantics based on description logics. DAML-Logic, that is expected to be released soon by the DAML community, will enhance the reasoning capability for ontologies represented in terms of DAML. In the meantime, description logic facilitates formalizing logical schema constraints as well as performing inference regarding their satisfiability.

DAML+OIL facilitates capturing data and service abstraction for complete SNE, but the advertisement of SNE on repositories simply requires mechanisms to formally specify the profiles and service (i.e., model interactions, major terrain data parameters) constraints. In the next subsection we argue on how these profiles could be utilized in SNE specification matching.

We use DAML-Services (DAML-S) language in our ontology broker design to perform specification matching during retrieval. DAML-S supplies service providers with a core set of markup language constructs for describing the properties and capabilities of their services in unambiguous, computer-interpretable form. DAML-S markup of services facilitates the automation of service tasks including automated service discovery, execution, interoperation, and composition. Following the layered approach to markup language development, the current version of DAML-S builds on top of DAML+OIL. In DAML+OIL, abstract categories of entities, events, etc. are defined in terms of classes and properties. DAML-S defines a set of classes and properties, specific to the description of services, within DAML+OIL. The class Service is at the top of the DAML-S ontology. Service properties at this level are very general. The upper ontology for services do not constrain what the particular subclasses of Service should be, or even the conceptual basis for structuring this taxonomy, but it is expected that the taxonomy will be structured according to functional and domain requirements. The ontology of services provides two essential types of knowledge about a service, characterized by the questions:

- *What does the service require of users, and provide to them?* This is provided by the profile, a class that describes the capabilities and parameters of the service (for both SNE data and model). More specifically, the profile contains the general service description, functional description, and functional attributes. Service description entails the name, intended purpose, and textual description of the service. The functionality description input, output, precondition, effect (postcondition), and invariant conditions on parameters. The functional attributes consider other aspects of the service such as geographic radius, degree of quality, service type and category, quantitative quality guarantees and ratings.

- *How does it work?* The answer to this question is given in the model, a class that describes the workflow and possible execution paths of the service (for SNE models). That is, a process ontology is provided to model the service as processes. Hence, the way the service is delivered can also be queried during environmental model selection.

### Web-Based SNE Component Retrieval with DAML-Service Advertisement and Discovery Protocol

The dynamic and autonomous ontology discovery ability is important to SNE users because ontologies will be continually evolving and new ones coming into existence. When a request to retrieve an ontology is received, the ontology (or only the relevant portions – if large) will be dynamically accessed and mapped to the reference ontology

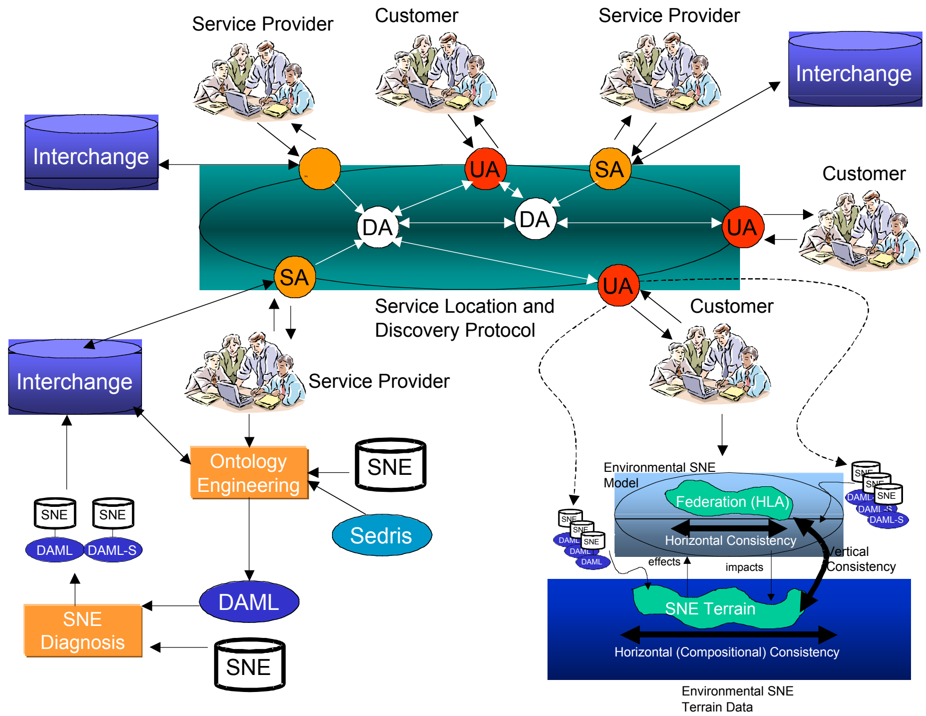


**Figure 2: Organizational Layout of SNE Service Publish, Discovery, and Retrieval**

***Principle Aspects of Service Advertisement and Location Protocol***

Matching the needs of user agents to the dynamically changing services offered by SNE service providers requires a service discovery protocol as shown in Figure 2. The components of the service location and discovery protocol and their functions are as follows:

(1) The advertisement of DAML-based service profiles by service provider agents.
(2) The organization of DAML service profiles into directories managed by directory agents.
(3) Obtaining service information for client (customer) agents from the Service Location Protocol (SLP) framework, without prior configuration.
(4) Providing a means for services to inform client applications of their capabilities and configuration requirements (optional).

Directory agents (DA) of the SLP, as shown in Figure 2, form a hierarchy, where each root component designates a certain facet to represent an independent aspect in the context of faceted classification as discussed in the previous section. For large repositories, a root directory agent contains a top-level index for terrain data schema and environment model metadata components based on facet classification.

Each lower level directory agent is either an index, a partial index and metadata advertisement data store, or a complete metadata store. The metadata stores constitute enumerated classification of DAML-S ontologies based on service profile categories. The service agents (SA) interface to service providers and the servers (ontology and SNE component repositories) that host the SNE terrain data and models. User agents (UA) contact service agents to identify and retrieve relevant SNE components along with the complete metadata stored in the ontology and component servers (Interchange server) hosted by the sites the service agents reside. They contact service agents based on the service access point information acquired from the directory agents as a result of successful *specification matching* operation during ontology querying process.

***Ontology and SNE Component Repository Server***

Here, we discuss how an object-oriented design repository, called *Interchange* with persistent DAML storage, retrieval, and manipulation API delivers the server requirements of the architecture depicted in Figure 2. The underlying design rationale for *Interchange* is that, typically, large projects with geographically distributed teams need an infrastructure that will provide an efficient data distribution mechanism.

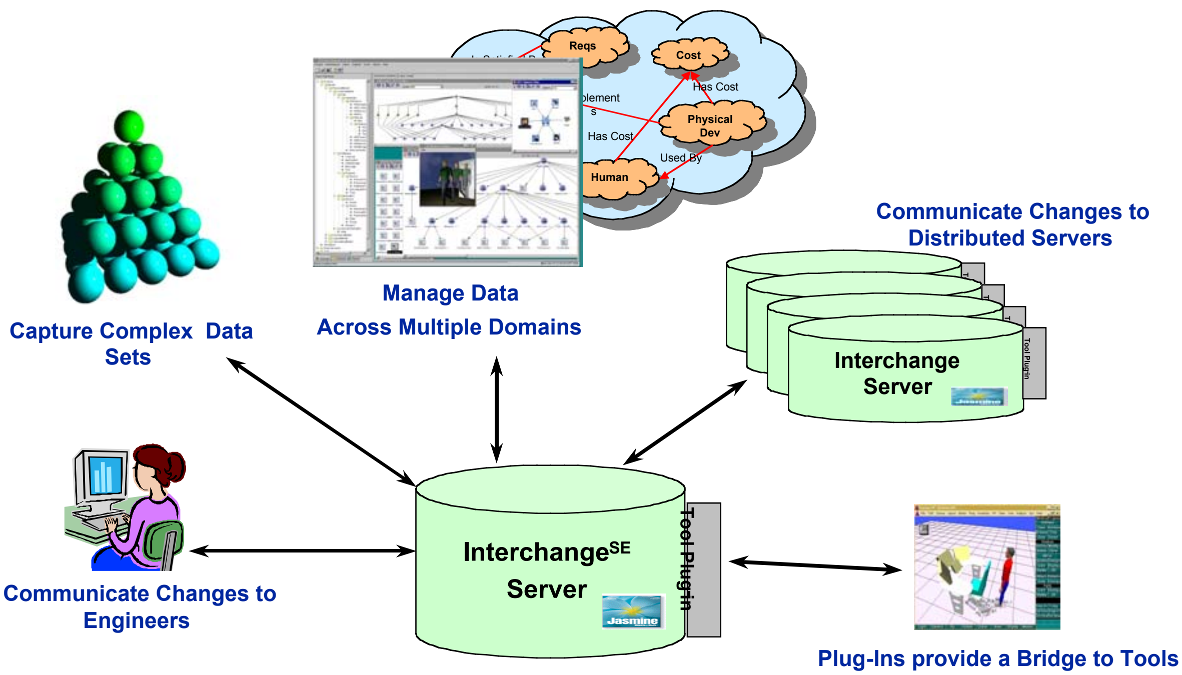


**Figure 3: Interchange Server**

The augmentation of Interchange design repository with the *service location protocol* that is briefly discussed above facilitates ontology construction, visualization, storage, retrieval, and manipulation for DAML and DAML-S metadata. For instance, upon a matching successful query, the directory agent forwards the service access point of the service agent that hosts the required SNE component whose profile has been successfully matched. The user agent contacts the identified service agent for complete SNE metadata (and optionally actual SNE data and model) delivery. The service agent uses DAML augmented Interchange repository API to retrieve the corresponding complete metadata and forwards it to the client (user agent).

Interchange$^{SE}$ provides Java based user interfaces to allow users to view and manipulate data. Users can create and save multiple workspaces that can contain various views of the data. There are four layouts for data in a view, Orthogonal, Symmetrical, Hierarchical, and Circular, which users can select. Interchange$^{SE}$ also provides a manual mode where the user can drag objects to their desired locations.

A DAML API and property editor would allow the user to modify attributes and set relationships between any DAML objects in the repository. The objects in Interchange$^{SE}$ can have embedded multimedia attributes. This media data can be viewed at any time with the Interchange$^{SE}$ Java media framework or with external viewers and applications such as Microsoft Word and Excel. Users can register their own file types and viewers with Interchange$^{SE}$.

There are several levels of Security, or Access Control, in Interchange$^{SE}$. There is basic project level (user names and passwords), object level (read and write permissions on objects) and attribute level access control (read and write permissions on specific attributes of an object). Communication between the Interchange$^{SE}$ client and server is encrypted using SSL and can work through firewalls. Interchange$^{SE}$ supports three types of internal agents. There are change notification agents that monitor objects for changes. When an object of interest is modified, email notification of the change is sent. A Query agent periodically executes a query and emails the results while a method agent periodically executes a method and emails the return value of the method. The person who creates the agent specifies who will receive email notification.

The agent manager allows users to activate, deactivate, view, modify or delete an existing agent. Interchange$^{SE}$ also provides configuration management of the data by providing object level versioning where history objects can be viewed, purged, deleted or reverted to.

These features provide a basis to support enable SNE repository administration. There is also a Query builder in Interchange$^{SE}$ that allows users to create, execute and store queries. The results of a query can be displayed in a view. Interchange$^{SE}$ has created an infrastructure that facilitates the sharing and manipulation of data from numerous engineering tools.

### *SNE Specification Matching Mechanisms*

The retrieval of relevant SNE components require a mechanism to formulate SNE specific queries and determine which published service profiles match the query constraints provided by the user. We have proposed two specification-matching schemes. The first one is the short-term structural (syntactic) solution that enables a multi-level specification matching mechanism, while the second one entails a semantic (content) solution using a new SNE feature-oriented DAML query mechanism with deductive capabilities.

Structural querying constitutes facet identification, profile identification, and ontology structural affinity measuring with schema similarity metrics. The three repository organizations mentioned above are synergistically combined as well to support SNE component identification.

*Facet Identification:* A syntactic keyword-based search helps identify the facet under which the SNE component is being searched. This search identifies the directory agent under which the profiles will be searched.

*Profile identification:* Every DAML-S profile provides service description, functional description, functional attributes, and process ontology (for models) along with domain ontology. Customer profile input is obtained from the client interface and a potentially relevant ontology is identified on the basis of (1) *service description* (service name, intended purpose, text description, service provider, (2) *functional description* for models (input, output, precondition, effect, domain resource, parameters, parameter restrictions, and condition specification that constitutes logical assertions denoting the preconditions and effects of model services, (3) *functional attributes* (geographic radius, degree of quality, service parameters, communication mechanisms, service type and category, quality guarantees, and quality ranking. Detailed preliminary service match can be performed based on the process ontology for models so that the way that service works (the process) can be queried as well.

*Measuring Structural Affinity with Similarity Metrics:* After profile matching is applied, ontology similarity test is performed to determine the degree of relevance of the relevant schemas with respect to required SNE federation (data or model) components based on their ontology concepts. The notion of SNE federation is discussed in the next section. Similarity between conceptual schemas is computed using their schema descriptors and applying an affinity criterion on the features contained therein. Intuitively, the greater the number of features that present affinity, the higher the similarity between the corresponding schemas. A *weighted sum* metric is studied to compute ontology similarities. However, other similarity metrics such as *dice function*, *fuzzy similarity,* or standard text similarity function, called *cosine function*, can also be used.

Keyword-based (including the usage of the synonyms) preliminary similarity matching enables retrieving potentially reusable components. However, the following method will enable further more accurate and precise formal specification matching as long as the customer (the client) provides DAML-Logic or descriptive logics statements denoting the required logical constraints on the domain ontology.

*Deductive Reasoning for Content (Semantic) Querying:* Description logic systems have been used for successfully building a variety of research prototypes including conceptual modeling, information integration, query mechanisms, view maintenance, software management systems, planning systems, configuration systems, and natural language understanding. In summary, Description Logics (DLs) is the most recent specification approach for a family of knowledge representation (KR) formalisms that represent the knowledge of an application domain that constitutes the relevant concepts of the domain and then using these properties of objects and individuals occurring in the domain. One of the characteristics of DL is that, it is equipped with a logic-based formal semantics. Another distinguished aspect is the emphasis on automated reasoning as a central service; that is, this aspect allows one to infer implicitly represented knowledge from the knowledge that is explicitly contained in the knowledge base.

We are currently working on DL-oriented SNE query language with sound semantics for SNE data and model query formulation. The next step is to develop the semantics (behavior) of the query language, develop a model of the SNE model base upon which the semantics will be defined, and finally implementation of the model. The approach aims at presenting SNE component retrieval as a unique discipline, endowed with its own goals and techniques.

## SNE Diagnosis and Compositional Verification using Descriptive Logics

SNE components are retrieved based on their metadata specification. This implicates the necessity for assuring that the metadata is correct and accurate representation of the actual data or model. Hence, SNE components need to be certified against their published specifications. Furthermore, assembling retrieved components to form a cohesive and consistent SNE requires compositional consistency analysis to detect conflicts. To this end, we have formulated a SNE diagnosis and compositional verification approach that targets these concerns.

### *SNE Diagnosis for Schema, Data, and Layout Conformance*

DAML has a well-defined axiomatic as well as model-based formal semantics. Most constructs in DAML are associated with logical axioms that can be utilized as the schema consistency properties. Checking the satisfiability of these axioms with respect to the actual SNE schema constitutes the schema conformance analysis component of SNE diagnosis. The data conformance analysis, on the other hand, utilizes property and property value restrictions on metadata specifications. A generic procedure that takes ontology specification and actual data as input to determine *property and layout restriction* violations in the data is devised. In particular, for property violations the ontology parser generates a constraint graph. The constraint graph is evaluated by the actual data. The leaf nodes in the graph denote the ontology property nodes denoting value restrictions, while the internal nodes denote the class nodes. The edges between class nodes reflect the constraints on the object associations. The generic procedure traverses the constraint graph and check if all constraints are satisfied by the data through scanning the actual data set. For instance if a class called "Tank" has a feature *x* with the property field that imposes a *range* restriction, for example, a number between 10 and 100, then any data component, which is not numeric or that has a property beyond the required range will be marked as a failed data. The feature layout conformance, on the other hand, utilizes *context-free graph grammar* formalism, to determine if all the entities in the actual data set are organized and related (association and aggregation relations) according to a graph grammar pattern specification.

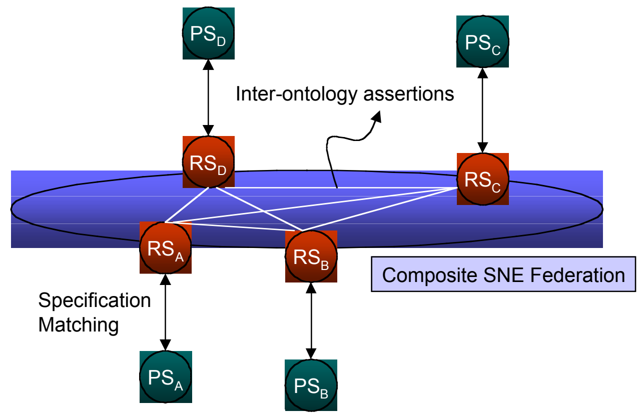


**Figure 4: The Consistency Notion for SNE Composition**

### *SNE Composition (Assembly) Consistency*

The compositional consistency analysis approach extends the intra-ontology consistency analysis point of view to inter-ontology domain. That is, domain independent as well as domain-specific axioms for inter-ontology relationships facilitates this analysis. The satisfiability of these assertions (axioms) reflects the consistency of the composition. We are in the process of investigating and adapting a descriptive logics based solution to asserting schema consistency relations. The goal is to axiomatize SNE domain entities and relations, facilitate inter-ontology integrity constraint definition using description logics, and perform reasoning to determine if these assertions are satisfiable by the assembly of retrieved SNE components.

The compositional consistency checks are based on the perspective of viewing composite SNE terrain data and models as a federation as depicted in Figure 4. In this figure *PS* denotes provided (reused/developed) service, while *RS* denotes required service. In our approach a composite SNE is originally specified and represented as a federation of SNE components. Each *RS* is an ontology with semantic constraints along with DAML+OIL. Specification matching methods discussed in the previous section are performed to identify potential relevant documents.

Note, however, even though individual SNE components (PS) locally match and fit to substitute corresponding required services, they may violate global constraints. Global constraints denote the inter-ontology constraints and will be defined in terms of DL assertions. Existing DL reasoners will be used to determine the satisfiability of these assertions.

## Conclusions

Our main position is that the retrieval of relevant SNE components in the context of a continuously changing dynamic information resources needs flexible, autonomous, intelligent publishing and retrieval mechanisms. Furthermore, the reusability of retrieved components in the application domain of interest requires identification of conflicts of the reused components with respect to the context they are embedded. Composition of retrieved SNE components and overlaying them on the terrain skin requires sophisticated schema-based consistency analysis and correlation methods that will augment the existing testbed. Our position is that performing consistency analysis at higher levels of SNE abstraction would facilitate avoiding costly data correlation until feasible and reusable components are identified with rapid abstract correlation mechanisms based on specification matching inference algorithms.

## References

DMSO (2000). "Integrated Natural Environment Authoritative Representation Process (INEARP) Concept of Operations", Technical Report.

Guttman Erik (1999). "Service Location Protocol: Automatic Discovery of IP Network Services", IEEE Internet Computing, July-August 1999.

Horrocks I., D. Fensel, F. van Harmelen, D. L. McGuinness, P. Schneider (2001). "OIL: An Ontology Infrastructure for Semantic Web", IEEE Intelligent Systems, March-April 2001.

Schaefer J., B. Graniela, P. Heard, J. De La Cruz (2002). "Web-Based Repository for STRICOM SNE Objects", Spring 2002 Simulation Interoperability Workshop.