<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>SESA: Emerging Technology for Service-Centric Environments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Vitvar, Tomas; Zaremba, Michal; Zaremba, Maciej; Moran, Matthew; Fensel, Dieter</td>
</tr>
<tr>
<td><strong>Publication Date</strong></td>
<td>2007</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td><strong>Link to publisher's version</strong></td>
<td>doi: doi:doi: 10.1109/MS.2007.178</td>
</tr>
<tr>
<td><strong>Item record</strong></td>
<td><a href="http://hdl.handle.net/10379/444">http://hdl.handle.net/10379/444</a></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
<td><a href="http://dx.doi.org/10.1109/MS.2007.178">http://dx.doi.org/10.1109/MS.2007.178</a></td>
</tr>
</tbody>
</table>
In large-scale, open, service-centric environments, thousands of services will have to be discovered, adapted, and orchestrated on the basis of user needs. Because SOA technologies employ only XML descriptions, they offer only manual support for integration that usually operates on a rigid configuration of workflows or services. Although XML is flexible and extensible, it defines only the data’s structure and syntax.

Extending SOAs with semantics offers scalable, adaptive automation. The goal is to design a semantically enabled SOA (SESA) and a technology promoting on-the-fly personalization and adaptability of business requirements. The SESA adopts a service model that uses semantics for rich description of both the services offered and the capabilities that potential users require. This model defines essential functionalities for the dynamic integration of services. The SESA aims to fulfill users’ goals through logical reasoning over semantic descriptions.

Ultimately, users won’t need to be concerned with the processing logic; they will be able to focus just on the result and its quality.

**Governing principles**

The SESA enables an open, service-centric environment where service orientation, intelligence, and seamless integration are the key to providing services to their users. Three principles drive SESA research, design, and implementation:

- The service-oriented principle promotes service reusability, loose coupling, abstraction, composability, autonomy, and discoverability.
- The semantic principle allows a rich description of information and behavioral models that enables automation through logical reasoning in service discovery, negotiation, selection, mediation, and so on.
- The problem-solving principle underpins...
the architecture’s ultimate objective: goal-based discovery and invocation of services. Users (service requesters) describe requests as goals semantically and independently from services, while the architecture solves those goals through logical reasoning over their descriptions, as we mentioned before.

The SESA and its related technology

The Web Service Modeling Ontology provides a conceptual model for describing domain ontologies, user goals, services, and mediators. To formally describe these elements, the Web Service Modeling Language defines several complementary language variants allowing for the logical expressiveness of both description logic and logic programming. WSMO and WSML provide the basis for a semantic technology that’s well suited for the SESA and its underlying principles. (For a comparison of SESA technology with other approaches, see the “Related Work on Adding Semantics to Web Services” sidebar).

SESA distinguishes two types of services. Middleware services are the main facilitators for searching, integrating, and mediating business services. The business services are exposed by the service providers’ back-end systems, which are integrated within the SESA. Through the functionality of these types of services, the SESA aims to support

- the business users who consume the functionality of business services through some domain applications, and
- the engineers (enterprise architects, application programmers, and domain experts) who consume middleware services to model, deploy, assemble, manage, and maintain business processes throughout the SOA life cycle.

As figure 1 illustrates, the SESA has three main layers: the middleware-services layer, the business-services layer, and the problem-solving layer. The WSMO conceptual model and its elements of ontologies, services, and goals map directly to the SESA business-services and problem-solving layers, while mediators map to the middleware-services layer.

The middleware-services layer

The middleware services reside in the semantic execution environment. They operate mainly on the semantic-service model of business services, aiming to facilitate the seamless integration of business services. The Oasis (Organization for the
Related Work on Adding Semantics to Web Services

Besides the Web Service Modeling Ontology (see the section “The SESA and its related technology” in the main article), the OWL-S\(^1\) ontology can be used for semantic description of services. OWL-S comprises three subontologies:

- **ServiceProfile** describes a Web service’s functionality.
- **ServiceModel** defines a Web service’s behavior.
- **ServiceGrounding** links ServiceModel and the description of the concrete Web service that the Web Services Description Language (WSDL) provides.

OWL-S uses the Web Ontology Language (OWL) for descriptions of its elements. OWL currently doesn’t support rule languages, which are essential for describing functional and behavioral service semantics using logical conditions. In addition, OWL-S doesn’t include the user request in its conceptual model, and it doesn’t assume that service environments are inherently heterogeneous.

To overcome these deficiencies, WSMO introduces a proper language layering, including logic-programming languages suitable for describing service semantics using logical conditions. It also introduces ontological role separation, clearly distinguishing between the requester and the provider side (that is, between goals and services). Plus, it introduces mediators as the core of a conceptual model treating heterogeneity as a natural aspect of Web services.

In addition, some technologies are building on the OWL-S model. Among others, the OWL-S Virtual Machine provides a general-purpose client for the invocation of Web services based on OWL-S descriptions.\(^2\) Although the OWL-S VM proves the OWL-S model’s value, it addresses only a subset of the functionality that a semantically enabled service-oriented architecture (SESA) offers. For example, the OWL-S VM doesn’t directly address mediation and lacks goal-driven execution.

Other specifications also aim to add semantics to services. The Semantic Annotations for WSDL and XML Schema (SAWSDL)\(^3\) is a recent World Wide Web Consortium (W3C) project that adopts WSDL-S\(^4\) specifications. SAWSDL aims to define simple extensions to annotate various WSDL elements with semantic descriptions. While SAWSDL doesn’t define any semantics specifically for service descriptions, it can serve as a grounding mechanism between WSDL elements and OWL-S elements.\(^5\)

Based on WSDL-S, METEOR-S (Managing End-to-End Operations with Semantics, http://lsdis.cs.uga.edu/projects/pastMETEOR) defines a complete service life cycle based on service semantics.\(^4\) It lets you create WSDL-S descriptions from annotated source code, automatically publish WSDL-S descriptions in enhanced UDDI registries, and generate OWL-S descriptions from WSDL-S for grounding.

In addition, you can view the SESA as complementing the W3C’s Web Services Architecture,\(^6\) building on its principles. However, some differences exist. For example, the SESA includes semantic mediation, refines the Web Services Architecture goal state to a WSMO goal, and includes WSMO choreography interfaces for both goals and services (a WSMO choreography describes a service’s external behavior). For conceptual models of Web services specifications (WS-*\(^7\)) such as policy, security, or reliability, the SESA either adopts these specifications as part of the business service model or adopts them through the underlying Web service technology.

References

The broker sublayer defines the functionality necessary for goal-based discovery and invocation:

- **Discovery** defines tasks for identifying and locating business services that can achieve a goal.
- **Adaptation** defines an adaptation within a particular integration process according to users’ requirements (for example, service contracting and validation).
- **Orchestration** defines the execution of a composite business process together with the communication between a service requester and a service provider in that process.
- **Monitoring** defines the monitoring of the business services’ execution. It gathers information on invoked services—for example, for determining quality of service or for identifying faults during execution.
- **Fault handling** defines how to handle faults occurring during the business services’ execution.
- **Mediation** defines interoperability at the data and process levels.
- **Composition** defines the composition of services into an executable workflow.
- **Grounding** defines transformations from semantic descriptions of business services to nonsemantic descriptions, and vice versa.

The base sublayer defines functionality that isn’t directly required in goal-based discovery and invocation but that the broker layer requires for successful operation:

- **Formal languages** define semantic languages used for semantic description of services, goals, and ontologies.
- **Reasoning** defines reasoning functionality over semantic descriptions.
- **Storage** defines a persistence mechanism for various elements (for example, repositories for services and ontologies).
- **Communication** defines the middleware’s inbound and outbound communication.

The business-services layer

The SESA adopts descriptions for business services using the WSMO service model while building on the underlying technology for invocation and communication. Here, we show how to define business services using descriptions of information, functional, behavioral, and non-functional semantics and how to ground those descriptions to WSDL message types and operations. (Although the semantic descriptions are independent of the underlying technology, the grounding to WSDL leverages existing de jure and de facto SOA standards in the SESA context.)

**Information semantics.** A business service’s core is the information semantics (information model) on which it operates. This semantics is the formal definition of the domain ontology that the service uses for description of functional and behavioral semantics.

In general, to model ontologies, you can use **concepts**, **attributes**, **relations**, and **axioms**. Concepts describe classes of objects and define the terminology of the domain of discourse. For example, in figure 2, the **Service** concept (line 6) stands for the class of all services that can be put in a subsumption relation by means of the **subConceptOf** construct (lines 10 and 11).
Attributes define relations between concepts, define these relations’ constraints, and point to data types. For example, in figure 2, the Customer concept has the attributes hasConnection and hasService. These attributes point to concepts for the parts of the Customer concept with the restrictions that the network connection definition is mandatory while at most one connection can be defined.

Axioms define arbitrary complex logical expressions over other definitions in the ontology. We show examples of logical expressions for some descriptions of services later in this article.

Functional semantics. This semantics is the formal description of service functionality, describing what a service can offer to its users when it’s invoked. We describe the service’s functionality as a capability, which we model as preconditions, assumptions, postconditions, and effects. As figure 3 illustrates, preconditions and assumptions define the conditions of the information space and of the world outside that space, respectively, that must hold for state $s_1$—the system’s state before the invocation of the service $W$. Similarly, postconditions and effects define the conditions for state $s_2$—the system’s state after the invocation of the service $W$. We then define the invocation of $W$ as a transition between states $s_1$ and $s_2$:

$$
\text{holds(PRECONDITIONS } \land \text{ ASSUMPTIONS, } s_1) \rightarrow \text{holds(POSTCONDITIONS } \land \text{ EFFECTS, } s_2).
$$

In figure 3, the state space comprises all the services available for a search (for example, services in a repository). The information space comprises all the information available to all the services. The world space comprises the information outside the information space. Here, the capability uses WSML to describe a network subscription service’s functionality—in this case, video on demand (VoD). The precondition specifies that the customer must have a connection with the minimal bandwidth the service requires. The assumption specifies that the customer must have sufficient funds in his or her bank account. The postcondition identifies a valid customer hav-

![Figure 3. A functional-semantics example.](image-url)
ing the connection and service defined. The effect specifies that the customer subscribes to the service.

**Behavioral semantics.** This semantics defines a service’s external and internal behavior. The external behavior, called a *choreography*, describes a protocol that a client uses to consume the service functionality. The choreography describes interactions involving messages sent to the service from the network and to the network from the service. (Our definition of choreography differs from the one that the Web Services Choreography Description Language uses.) The internal behavior, called an *orchestration*, describes a workflow—that is, how the service’s functionality is aggregated out of other services.

To model the choreography and orchestration, we use an *abstract state machine*. We define the ASM as

\[ I_W = (\text{In}, \text{Out}, L), \]

where \( I_W \) is the interface of the choreography or orchestration, \text{In} and \text{Out} are ontological concepts of some information semantics used in input and output messages, and \( L \) is a set of rules. Each rule has the form “if *condition* then *effect*” and corresponds to one or more interactions with the service following the underlying *message exchange pattern* (we explain MEPs in more detail later). For the rule to execute, the condition must hold in a state in the information space; the effect defines how the state changes when the rule executes—that is, how the information space is modified by adding, deleting, or updating the data.

Figure 4 uses two rules to define the choreography for the service in figure 3. The first rule specifies how to get the details about the customer (lines 15–19). This rule can execute if \text{SearchCustomerRequest} is
in the information space (in our example, we assume that the user will supply this information, as we explain later). The second rule depends on the first rule and defines how to perform the service subscription (lines 21–26). Here, the subscription request must use the customer’s identifier (customerId).

Nonfunctional semantics. This semantics describes additional constraints over service functionality that the functional description didn’t capture. They’re incidental details specific to a service’s implementation or running environment, independent of the service’s actual purpose but necessary for successful communication. To express these constraints, we can use policy languages (for example, WS-Policy) and their semantic representation. Nonfunctional descriptions can also capture other information about services, such as the author or version, using the Dublin Core metadata set (http://dublincore.org). In addition, we use them for grounding, as we describe next.

Grounding. The SESA business services adopt the WSMO semantics for various descriptive parts of services. For the parts regarding service invocation (such as how and where the client can access the service), the WSMO defines grounding from the semantic descriptions to the WSDL descriptions. The SESA uses this grounding for on-the-wire message serialization (WSDL binding), physical Web service access (WSDL service and end point), and communication (SOAP).

In addition, business service descriptions don’t preclude the use of other relevant Web services specifications (WS-* such as policy, service security, reliable messaging, and transaction). While policy specifications can have semantic representations as nonfunctional descriptions, the other specifications can serve as part of the WSDL descriptions. To support such specifications at the SESA middleware layer, relevant middleware services should implement them.

The grounding specifies

- references for ontological concepts used in a choreography to WSDL messages and
- transformations from XML data to instances of ontological concepts (lifting and vice versa (lowering)).

The grounding is useful during communication with the service after the definition of WSDL operations and their MEPs.

Table 1 shows four basic types of WSDL 2.0 MEPs (www.w3.org/TR/wsdl20-adjuncts/#meps) and their corresponding choreography rules and WSDL operations. Here, c1, ..., c6 are ontological concepts specified as input (In) or output (Out) of some choreography <I>, msg1, ..., msg6 are XML Schema elements of input or output messages of operations.

<table>
<thead>
<tr>
<th>MEP and rule</th>
<th>WSDL operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>in-out:</td>
<td>&lt;operation name=&quot;oper1&quot; pattern=&quot;http://www.w3.org/ns/wsdl/in-out&quot;&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;input messageLabel=&quot;In&quot; element=&quot;msg1&quot; /&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;output messageLabel=&quot;Out&quot; element=&quot;msg2&quot;/&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;/operation&gt;</td>
</tr>
<tr>
<td>in-only:</td>
<td>&lt;operation name=&quot;oper2&quot; pattern=&quot;http://www.w3.org/ns/wsdl/in-only&quot;&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;input messageLabel=&quot;In&quot; element=&quot;msg3&quot; /&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;/operation&gt;</td>
</tr>
<tr>
<td>out-only:</td>
<td>&lt;operation name=&quot;oper3&quot; pattern=&quot;http://www.w3.org/ns/wsdl/out-only&quot;&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;output messageLabel=&quot;Out&quot; element=&quot;msg4&quot; /&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;/operation&gt;</td>
</tr>
<tr>
<td>out-in:</td>
<td>&lt;operation name=&quot;oper4&quot; pattern=&quot;http://www.w3.org/ns/wsdl/out-in&quot;&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;output messageLabel=&quot;Out&quot; element=&quot;msg5&quot; /&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;input messageLabel=&quot;In&quot; element=&quot;msg6&quot; /&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;/operation&gt;</td>
</tr>
</tbody>
</table>
ref(c, m) denotes a reference grounding of an ontological concept c and a message msg.

As figure 5 shows, more complex rules usually exist whose conditions define the order in which the invocation should happen. The choreography specification's current version doesn't handle other types of MEPs, such as ones including faults and optional messages.

Figure 4 shows the example for the reference grounding (lines 5, 7, 10, and 12). A lifting or lowering grounding defines a transformation (usually in XSLT, Extensible Stylesheet Language Transformations) that's specified as a URI (uniform resource identifier) in the lifting or lowering nonfunctional description of the service (not shown in figure 4). When the rule for obtaining customer details (lines 15–19) executes (the condition holds in the information space), the system first lowers the instance data of SearchCustomerRequest to the XML message referenced by the concept (line 5). The system then invokes the corresponding WSDL operation, passing the XML message as the input data. Next, the system lifts the output XML message received from the invocation to the instance of the referenced Customer concept (line 10). Finally, the system adds the resulting data to
the information space according to the rule’s action (line 18).

**The problem-solving layer**

Through this layer, users can formulate or identify goals, submit goals, interact with the architecture during processing, and get desired results. End users can perform these activities through some domain applications; engineers can perform them through some management tools—that is, an integrated development environment. The reference implementations of the IDE framework developed for the SESA are the Web Service Modeling Toolkit (WSMT) and WSMO studio (www.wsmostudio.org).

The key to this layer is the formal description of the user’s objectives. For this purpose, the SESA defines the goal description as a requested capability (what the user wants to achieve) and a choreography interface (how the user wants to communicate) having the same structural definition as the business service (see the sections “Functional semantics” and “Behavioral semantics”). In addition, we can express user requirements such as a service’s desired quality by using nonfunctional descriptions of the goal.

Figure 5 shows a goal of a user who wants to subscribe to a VoD service. The precondition (lines 5–7) specifies that the user is a customer already having some network connection. The postcondition (lines 8–11) specifies that the customer subscribes to the intended VoD service. The choreography interface defines which data the user expects to supply (specified in `OrderRequest`) and to receive (specified in `OrderResponse`). The rules in lines 23–33 correspond to the out-in MEP of the underlying WSDL (see table 1). Here, `naf` (negation as failure) is a nonmonotonic inference rule that derives `not OrderRequest` when `OrderRequest` fails to be derived in the information space.

**Integration**

The SEE middleware defines two phases for business services integration: **late binding** and **execution**.

**Late binding**

This phase allows binding a user goal and a set of business services by means of intelligence in the SEE middleware. We call this phase “late binding” because the binding isn’t known a priori (that is, during modeling) and can be performed in a semiautomated way on the fly. Typically, late binding involves tasks for service discovery, contracting, and validation, complemented by composition and interspersed with mediation. This phase operates on semantic descriptions of business services and can take several forms, depending on the scenario.

**Discovery.** When the system receives a user goal, the discovery task matches the goal capability with the capabilities of services in the middleware repository. Taking the goal from figure 5 as an example, this task finds a potential service satisfying the goal capability in figure 3 (the goal postcondition concept `VideoOnDemandService` subsumes the `Service` concept from the service postcondition). The discovery task must evaluate various semantic relationships between the goal and services, including exact match, intersection match, subsumption match, plug-in match, and disjointedness.

**Contracting.** This process refines the discovery results by consulting dynamic instance data of the goal and the potential service. Including such data in static service descriptions is clearly infeasible—the system must obtain the data dynamically during the late binding.

In our example, the goal data is part of the service request as shown in figure 5 (through the rule in lines 23–27), while the system obtains the service data by executing appropriate rules from the service choreography in figure 4 (that is, it obtains the customer details using the rule in lines 15–19). In this example, we assume that through the order request in figure 5, the user supplies the customer identification together with the details for the subscription request needed by the service choreography in figure 4. If the goal and service choreographies use different ontologies, data mapping and mediation are necessary. The potential service can satisfy the goal if all data is fetched and the service precondition evaluates to true in the information space. In our example, the precondition holds if the customer’s connection bandwidth is greater than the service’s required connection bandwidth. More information about this process appears elsewhere.

**Composition.** When no single service can satisfy the whole goal, the composition task tries to create a plan for that goal. Using AI methods such
as backward or forward chaining, this task tries to find an order of services while recursively performing discovery and contracting. This results in a description of a composite business service's orchestration.

In our example, if the customer connection bandwidth is less than the service connection bandwidth, composition might result in a workflow of two services—that is, an upgrade connection service and the subscription service.

**Validation.** Although late binding aims to automate integration, human validation might be necessary in sensitive domains such as e-health, where late-binding results could have crucial real-world effects. In the SESA, a combination of service-analysis and user-approval methods ensures this validation.

Service analysis provides information based on the service usage and quality, answering questions such as “how often does the service fail?” and “how often is the service used in the given context?” With the help of this information, when the system notifies the user that the goal can’t be fully satisfied, he or she can relax some of the functional or nonfunctional requirements. By selecting less functionality, the user might find higher-quality services; by relaxing the nonfunctional requirements, the user might find services with more functionality.

**Execution**

Figure 6 depicts this phase, which is performed by three middleware services. Orchestration manages the whole execution process, mediation resolves heterogeneity issues, and communication implements grounding together with inbound and outbound communication.

The process involves nine possible states:

1. The execution starts when the system receives a user goal (representing a requester) and a service (representing a provider). These descriptions are usually the result of the late binding. We also assume the software engineer has created the mapping of the goal and service ontologies and stored it with the middleware.

2. The system loads the goal and service choreographies into the orchestration engine, creating one instance for each. Each instance contains processing memory (representing the information space) used for storing data for runtime processing and a set of rules from the choreography definition.

3. The execution waits for new data (state 4),
processes the available data (states 5–8), or ends the execution (state 9).

4. The system checks whether new data from the user or service (both follow their choreographies) is available. If the system receives new data, it lifts the data using particular grounding definitions.

5. When new data is available, the system passes it to the data mediator. This mediator mediates the data from the user to the service ontology or from the service to the user’s ontology.

6. Control then passes to the process mediator, which decides where to put the new data: into the processing memory of the user’s choreography, the processing memory of the service’s choreography, or both. This mediator bases its decision on evaluation of the rules’ heads for each choreography. In particular, the mediator evaluates whether the subsequent processing of the choreography can use this data.

7. The system processes the updated choreographies. That is, for a rule whose head satisfies the content of the choreography’s processing memory, the system executes the body.

8. For each input concept in the rule’s condition, the system sends the data to the service or user, according to the definition of the underlying operation. To do this, the system lowers the data to the corresponding XML message. The execution then goes back to the control state (state 3).

9. When no data is available for the processing, no rules remain to be processed in choreographies, and no data remains to be received from the user or the service, the execution reaches the end state.

W hile the SESA facilitates a novel style of integration of services by means of semantic service descriptions and AI methods, some people say that such an approach isn’t realistic today. They argue that the complexity of semantic languages and integration techniques that depend on logical reasoning is a burden for service processing and high performance.

However, the logical reasoning can efficiently help resolve inconsistencies in service descriptions as well as maintain interoperability when these descriptions change. The more complex the services’ descriptions are, the more difficult it is for a human to manually maintain the integration. The semantics that promote the automation is the key to such integration’s flexibility and reliability.

To demonstrate the value of semantics for service descriptions as well as automation in service integration, we’re working on the Semantic Web Services Challenge (www.sws-challenge.org). The SWS Challenge aims to establish a
common understanding, evaluation scheme, and testbed to compare and classify various approaches to services integration in terms of their abilities as well as their shortcomings in real-world settings. Although a world full of services doesn’t exist yet, one-click integration will be desirable. The SESA and its related activities enable such a world as well as such integration.

Discuss this article at http://blog.vitvar.com/SESA.

Acknowledgments
Science Foundation Ireland grant SFI/02/CE1/I131 and the EU projects Knowledge Web (FP6-507482), SemanticGov (FP6-027517), and SUPER (Semantics Utilised for Process Management within and between Enterprises, FP6-026850) supported this research. We thank Jacek Kopecky for his input on nonsemantic descriptions of services.

References