6.4 - Tie-in with Terminology and Platform

This section describes the operational dependencies between the semantic infrastructure and terminology and the platform and includes the following:

Dependencies between Semantic Infrastructure 2.0 and caGrid 2.0

Refer to the following sections of 6 - Dependencies Between Semantic Infrastructure 2.0 and caGrid in the caGrid 2.0 Roadmap Documents

- Semantic Infrastructure Registry
    - Registry Reliance on Platform
    - Platform Reliance on Registry
- Metamodel and Information Model

Semantic Infrastructure Overview

In an effort similar to developing this roadmap for Semantic Infrastructure 2,0, a team is developing a roadmap for the future platform, security and tools, caGrid 2.0. The Semantic Infrastructure 2.0 will be tightly integrated with the runtime caGrid 2.0. The purpose is to achieve a more comprehensive approach to computable semantic interoperability then is possible with the existing integration between caDSR and caGrid 1.x. With the adoption of SAIF (Service-Aware Interoperability Framework) and Enterprise Conformance and Compliance Framework (ECCF) and the introduction of behavioral semantics, the infrastructure of the grid must provide increasingly sophisticated support to leverage and enforce behavioral specifications.

The notion of "computable semantic interoperability" (CSI) applies semantics to not only the static data passed between machines, but also to the behavioral and functional operations exposed for coordination of behaviors during the interaction. Likewise, the semantic infrastructure itself (that is, its tools and applications) are being transformed to fully participate in the new services-aware environment. Thus the Semantic Infrastructure will depend on the grid platform as at least one of potentially many delivery platforms for its information.

The Semantic Infrastructure 2.0 is expected to provide management services and tooling comparable to those which exist today (including vocabulary services, model management and annotation, and curation tools and others as needed), albeit in potentially new formats and standards. However, it also has the increased scope of greater flexibility in accommodating granular levels of conformance and participant sophistication.

For example, the assumption of adherence to a centrally curated authoritative source of information models and terminology is no longer true; the infrastructure must gracefully accommodate local terminologies or localizations as well as standard terminologies. It must enable a path towards "as much interoperability as is possible" between any two parties, rather than enforcing full "compatibility" of all participants. Additionally, the Semantic Infrastructure 2.0 is charged with making the wealth of knowledge contained within the numerous SAIF artifacts available and consumable (in a programmatic fashion) to all the grid participants. Having runtime support for leveraging this information to inform and drive service interactions is a key value proposition for the future platform and semantic infrastructure.

Semantic Infrastructure Registry

The key components of the Semantic Infrastructure 2.0 are still being identified and scoped. The Semantic Infrastructure Registry has been identified as a key component. It is expected to be critical to the grid. The Semantic Infrastructure Registry may ultimately be manifested as numerous types of registries and services; potentially there may be numerous instances of each.

> ⓘ **Note**
>
> The ECCF registry provides storage for the semantic components of ECCF artifacts as specified by governance. The ECCF Registry will be populated by the iterative process of service design and specification.

The specification of a CBIIT enterprise service specification requires the development of three separate artifacts:

- The CIM (computationally independent model specification)
- The PIM (the platform independent model specification)
- The PSM (the platforms specific model) specifications

The CIM, PIM and PSM are again a collection of artifacts (models). The ECCF matrix is placeholder for these artifacts organized by Reference Model of Open Distributed Processing (RM-ODP) viewpoints and model-drive architecture (MDA) perspectives.

Currently a Microsoft Word document acts as a template or placeholder for describing the Computation-Independent Mode (CIM), Platform Independent Model (PIM) and Platform Specific Model (PSM) (along with the artifacts of each viewpoint), while the future semantic infrastructure will define computable representation formats for this information.

Part of this computable representation is expected to be a Service-Oriented Architecture (SOA) ontology for describing various entities involved in the numerous conformance assertions (with examples including but not limited to services, operations, data types, faults, and actors). This ontology will provide the backbone for reasoning to be performed by the platform and tools at both runtime and design time (as illustrated later in this section).

Registry Reliance on Platform

While the Semantic Infrastructure Registry services will be specified in ECCF, and potentially manifested on multiple platforms, one such platform will be the grid and will therefore use the platform as scoped in this document.

The registry will require numerous capabilities described in this document including the security layer for items such as authentication, authorization, auditing, and data assertions and integrity. The infrastructure may also require support for a rules engine capable of consuming and enforcing rules in both static (for example, data constraints in the information model) and behavioral semantics (for example, pre-conditions and post-conditions of operation invocation) stored in the registry.

Similarly, the platform will inform the content and format of various platform specific artifacts to be stored in the registry (including but not limited to XML Schema (XSD) and Web Service Definition Language (WSDL)). The platform will provide the capability to enforce or test conformance to those profiles by, for example, checking service interfaces against published PSMs and doing data validation against published information models. Finally, the platform will act as the service implementation technology for the grid services of the registry (that is, be the management interfaces or consumer facing services).

Platform Reliance on Registry

The platform itself will require and leverage numerous capabilities of the Semantic Infrastructure 2.0, most importantly, access to the information contained in the ECCF registry. The registry will facilitate nearly all parts of the service development and consumption life cycle.

At design time, the registry will provide a wealth of information to the designer including available relevant service specifications to adopt and extend, information models and terminologies to leverage for new operations, and formal specifications of expected behavior of the existing services that the new service may consume from in its implementation. For example, service templates (shelled out implementation artifacts) could automatically be constructed based on platform specific specifications.

Extending beyond the basic query and retrieval of these artifacts, tools can be built to actually understand the semantics of this information and aid the service developer. For example, potentially relevant information models may be found by entering simple terms like "tissue sample" into a tool, which binds that string terms to concepts in identified terminologies and locates models containing information bound to those terms. Similarly, behavioral contracts may be discovered based on terminology binding to their function based on simple search terms like "data insert," which can act as models or examples for new service operations.

Further, such "understanding" of behavioral and static semantics can provide a powerful feature in a tool for workflow or service composition. It could leverage this information to make suggestions on "next steps" in a workflow, even suggesting specific services to use. It could also provide powerful integrity checking of the data flow and functional effect, by validating that the invocations are consistent with published conformance assertions and rules.

At deployment time, the platform (or deployment tools built on it) can automate the generation and execution of a test suite to check conformance assertions published in the registry, relevant to the service being deployed.

At run time, the service can provide powerful self-descriptive metadata by referencing profiles, policies, conformance assertions, and specifications in the ECCF registry. This metadata will provide significant details about the nature and behavior of the service, and can be used to discover it, as well as to ascertain programatically how to correctly consume it (and validate it is functioning correctly). The platform may also be able to automatically flag non-comforming service instances (for example, services sending incorrect data, or running outside of published performance metrics) by monitoring runtime behavior.

Metamodel and Information Model

The Semantic Infrastructure 2.0 effort is still deciding on the format and structure to be used. This decision will be important to caGrid 2.0, as it will inform how things like the publishing of service metadata work, and how higher layer semantics (for example, operation preconditions) are built upon static descriptions (for example, WSDL). It is expected, however, that a transition from ISO 11179 metadata to RIM-derived semantics is important in the future infrastructure. As further information is available from both roadmap efforts, this section will evolve to discuss the impact on the platform.