

7 - Gap Assessment for Semantic Infrastructure

Current Working Draft

The section provides an assessment of the gap between the roadmap and existing tools and platform. The following topics are included:

- [Existing NCI Semantic Infrastructure](#)
- [Proposed Features in Semantic Infrastructure 2.0](#)

Existing NCI Semantic Infrastructure

The NCI semantic infrastructure currently consists of a suite of tools aimed at terminology curation of models submitted as UML XMI files for semi-automated annotation; terminology services for concept lookup and codesystem browsing; and basic terminology and ontological relationships in the NCI Thesaurus and Metathesaurus. This bundle of infrastructure applications together with model-driven software engineering tools are termed caCORE (Cancer Common Ontologic Representation Environment).

caCORE tools and APIs are developed by the National Cancer Institute Center for Bioinformatics and Information Technology (NCI CBIIT) to provide the building blocks for development of interoperable information management systems. This suite of tools has helped to enable interoperability and data sharing from the scientific bench to the clinical bedside and back with the current semantic infrastructure.

caCORE includes the following key components:

- [EVS](#) (Enterprise Vocabulary Services) for hosting and managing vocabulary
- [caDSR](#) (Cancer Data Standards Registry and Repository) for hosting and managing metadata
- [caCORE SDK](#), the GUI-based [caCORE Workbench](#), and associated tools for model-driven software engineering of systems which can be easily integrated with [caGrid](#).

[EVS](#) and the [caDSR database and tools](#) are the current basis of the semantic foundation for interoperable data and analytical services at NCI. caDSR is based on the ISO 11179 Part 3 metadata standard.

Developers use caCORE components to create "caCORE-like" systems. By definition these systems have object-oriented information models registered in caDSR whose meaning is linked to EVS vocabularies, and have open, public APIs and web services to provide access to the data. The [caBIO data service](#) is an example of a caCORE-like system developed using caCORE components.

Using caCORE tools, developers adapt and build applications that are [caBIG® compatible](#), that is, interoperable with other caBIG® tools.

caCORE tools include the following:

- [caDSR APIs Download](#)
- [CDE Browser; DTDs](#)
- [Form Builder](#)
- [CDE Curation Tool](#)
- [caDSR Administration Tool](#)
- [UML Model Browser](#)
- [Semantic Integration Workbench](#)
- [caDSR Sentinel Tool](#)
- [NCIThesaurus](#)
- [NCIMetathesaurus](#)

Additionally caCORE includes the [caCORE workbench](#), a tool with a graphical user interface (GUI) to facilitate the creation of a caBIG® silver or gold compliant system. The caCORE Workbench acts as a process guide and an integrated platform, enabling the user to more readily create a Data or Analytical service on the Grid. The following caBIG® process workflows are supported:

- Creation of a UML Model (ArgoUML, Enterprise Architect)
- Semantic integration (SIW, CDE Browser, UML Model Browser, Curation Tool)
- Model mapping (caAdapter)
- Application creation and deployment (SDK)
- Creation of a grid service (Introduce)

Proposed Features in Semantic Infrastructure 2.0

Semantic Infrastructure 2.0 is meant to provide a means of fully supporting the existing NCI semantic infrastructure, while providing a means for ongoing transformation of the existing artifacts and creation of equivalent tooling to support all current functionality of the semantic infrastructure.

Semantic Infrastructure 2.0 extends the current functionality of the semantic infrastructure by adding the following functionality:

- A new means of assessing conformance of artifacts and applications to improve software development and semantic consistency
- A semantically linked artifact repository for easy discovery of the registry contents
- A metadata repository that links to the artifact repository
- A cross-artifacts editing dashboard that allows model artifacts to be linked to other artifacts such as terminology value sets
- A rules engine for operating on the artifact repository and metadata repository to enable dynamic annotation and the comparison of artifacts
- A reasoning platform that executes inferencing and links to rule engines enabling the discovery of implicit information rather than explicit information only

- Introduction of additional semantic modeling standards (ISO 21090, HL7 Reference Information Model (RIM), Semantic Web Languages (Web Ontology Language (OWL), Resource Description Framework (RDF)) in order to handle the broad requirements of enabling simpler query functions and enriched data discovery
- A more automated artifact governance platform that includes the ability for community input to governance decisions
- Multiple model transformation tools and APIs
- Tools for authoring standards-compliant artifacts including schemas, models, and terminology value sets
- Tools for authoring forms using the new semantic models in order to meet the demands of customers who require these so that they can meet meaningful use requirements, and who want full semantics for data aggregation and discovery
- Broad use of Model Driven Architecture technologies
- Close integration with caGRID 2.0

The table below shows a high level view of the gaps between what the current semantic infrastructure provides and what Semantic Infrastructure 2.0 will provide for several use case-driven functionalities.

Requirement	Current Semantic Infrastructure	Semantic Infrastructure 2.0	Gap closed
Retrieve any artifact	CDE	Domain models, Logical models, terminology, documents, forms, behavioral models an specifications	Ability to retrieve any artifact in context
Manage artifacts	CDE Curator only	Open to all	Ability for anyone to annotate an artifact and submit to governance
Service discovery	Constrained to service discovery on caGrid	Service discovery tied to artifacts that can link to data provision	Ability to discover a service, its links to other services, the service contract, the artifacts that are behind the service
Bench to Bedside Form creation	Clinical research form creation	Form creation of any healthcare, clinical research, or life science form	Supports all form users and conforms to Office of the National Coordinator (ONC) requirements for meaningful use forms
Decision support across artifacts	None	Semantic linkage across multiple artifacts, inference of implicit knowledge about the artifacts and their relations	Provides enhanced search and retrieval of artifacts and extends the metadata for any artifact through inference of relations
Conformance Testing	None	Semantic reasoning and inference with automated classification, relations, and traceability of artifacts	Provides the full traceability and conformance testing for artifacts in a standard framework (Enterprise Conformance and Compliance Framework (ECCF))
Data discovery	Able to query caDSR for a model attribute and return an attribute identifier and reuse that identifier in a query for data	Semantic inference, semantic to relational adapters and scalable relation graphs relate services to artifacts, artifacts to terminology, and terminology to data allowing queries of models, classes, concepts or any other artifact and its data	Provides the ability to link services to each other and to the explicit definitions of the data they provide