

# Semantic Infrastructure Concept of Operations - Existing caBIG Semantics Implementation

## Semantic Infrastructure Concept of Operations - Existing caBIG® Semantics Implementation

### Semantic Infrastructure Concept of Operations Links

- [Semantic Infrastructure](#)
- [Vision](#)
- [Background](#)
- [Mission](#)
- [Objectives](#)
- [Overview](#)
- [Initiatives](#)
- [Stakeholders](#)
- [Alternatives](#)

Today's caBIG® Semantic Infrastructure forms the basis of the semantic interoperability capabilities embodied in the implementation of Data Services on the caGrid.

The existing caBIG® semantics implementation supports only the recording of data semantics. It does not embody all various properties of the three viewpoints shown in the [VKC:logical components diagram](#), the [VKC:contract design diagram](#) and the [VKC:deployment runtime diagram](#) (explained in discussion of architectural infrastructure in these pages) and required to support services discovery and reuse.

Specifically:

- The role played by each system is implicit. In a services paradigm, this is sufficient for the client, but not for the service implementation. The role should be characterized by standards in accordance with RM-ODP viewpoints and describe the policies including permissions, obligations, and prohibitions.
- The data service has a single operation: Query. This service has a contract that dictates the terms by which it may interact with other systems, "Query by Example" (QBE), and is intended for use with any information type. It may not have rules associated with the information it passes through the service.



#### Additional Information

The pre-condition behavior of a contract is issue a request to the index service for an end point reference. The behavior is the fulfillment of the query. The contract is between the client (initiator of the query) and the system (the responder to the query). The contractual context exists for the period of the behavior.

- In contrast, the information that is exchanged through analytic services may be built from an analytical process – it may be necessary for the rules to be modeled with the information.

The compatibility guidelines serve to insure that the behavior can be performed and, separately, that the objects exposed are sufficiently annotated semantically. However, since these semantics are at least partially derived from rules about the information, depending on how the information was modeled, there is a potential disjoint between in the details, available through a general query interface, and the requirements needed for usage of the information.

The disjoint between the specified role, the interface, the information objects, and the environmental contract of the deployed service implementations seems not to be a problem in areas of high trust where business rules are not associated with information objects or where their application is unimportant. This is true in many of the research applications.

Where this same pattern is intended to apply to more business-oriented implementations (clinical data), this pattern is difficult to implement because of the disjoint between the rules and the exposed information. In other words, when a developer wants to simply expose data through a Data Service, this pattern may be sufficient and powerful, but when it is intended to be repurposed for other uses, problems may arise. For example, it is typical for a research data service to expose the data points that resulted from an analysis "Smoking Indicator," but not to expose the set of raw data that was analyzed in order to make this assessment. This limits the usefulness of the data point "Smoking Indicator" since the consumer of the data wouldn't know what criteria were used to make establish this indicator's value.

caBIG® already provides a vast framework for tools, tooling, and other pieces of infrastructure. Using the architectural infrastructure noted above, the existing caCORE tools and some of the caGrid infrastructure may be categorized by the component of the new Semantic Infrastructure to which they relate. Re-use and adaption of the existing semantics is preferred over replacement whenever that is the cost effective solution to semantic requirements.

Below is an initial list (perhaps partial) of existing tools categorized by the project or package in which they are currently developed and managed. This list will serve as a starting point for the comprehensive mapping and assessment of the gaps between current infrastructure and the architectural infrastructure noted above, which are intended to support a caBIG® enterprise SOA. While existing tools may only partially meet the functionality of the proposed architecture infrastructure, the rationale for using, adapting or replacing them will be contained in the project documents generated for each of the initiatives proposed by this Concept of Operations.

### Initial Allocation of Existing Tool Packages to Components of the new Semantic Infrastructure Packages Component

Packages	Component
caAdapter	<p>Provides model mapping services in support of building caCORE-like applications and facilitates data mapping and transformation among different kinds of data sources including:</p> <ul style="list-style-type: none"> <li>• HL7 v2 messages</li> <li>• HL7 v3 messages</li> <li>• SDTM (Study Data Tabulation Model) data sets</li> <li>• MMS (model mapping service, object to relational mapping)</li> <li>• GME Mapping (create mappings between schemas) Mapping Tool Core Engine</li> </ul>
caDSR	<p>A collection of tools and services that support 11179 metadata development, registration and access. See individual tools for product descriptions.</p> <ul style="list-style-type: none"> <li>• UML Loader</li> <li>• CDE Browser</li> <li>• Admin Tool</li> <li>• Curation Tool</li> <li>• Form Builder</li> <li>• Sentinel Tool</li> <li>• Freestyle API</li> <li>• caDSR Services</li> <li>• Bulk Loader</li> <li>• Semantic Integration Workbench</li> <li>• UML Model Browser</li> <li>• XMI Handler</li> <li>• Object Cart</li> <li>• Domain Class Browser</li> </ul>
caGrid	<ul style="list-style-type: none"> <li>• Global Model Exchange (GME)</li> <li>• Index Service</li> <li>• Metadata Model Service</li> <li>• Grid Trust Service</li> <li>• Grid Grouper</li> <li>• Dorian</li> <li>• Credential Delegation Service</li> <li>• Authentication Service</li> <li>• Federated Query Processor</li> <li>• Taverna Workflow</li> <li>• BPEL Workflow</li> <li>• Introduce</li> <li>• Data Services</li> <li>• Web Single Sign On (Web SSO)</li> <li>• Transfer Service</li> </ul>
CDE Browser	<p>"Data Element Centric" Browser – view of caDSR content from perspective of individual data elements – organized by owning Context and classifications within each context, including folders containing the CDEs that are associated with data collection forms that have been registered in caDSR.</p>
CDE Curation Tool	<p>A 11179 content development tool that aids in the development and maintenance of caDSR metadata.</p>
Common Security Module (CSM)	<p>A solution to allow application developers to integrate security into their services with minimal coding effort. It is integrated with caGrid security framework and helps eliminate the need for development teams to create their own security methodology.</p>
Enterprise Vocabulary Service	<p>A collection of tools and services that support terminology development, maintenance and access. See individual tools for product descriptions.</p> <ul style="list-style-type: none"> <li>• LexEVS</li> <li>• Bioportal (retired June 2010)</li> <li>• NCI Term Browser</li> <li>• NCI Report Writer</li> <li>• NCI Protégé</li> <li>• Semantic Media Wiki</li> </ul>

<a href="#">Form Builder</a>	A caDSR tool for developing, maintaining and accessing metadata descriptions of CDEs grouped into modules as a series of question/answer sequences that together create the metadata representation of a "form".
<a href="#">Semantic Integration Workbench</a>	A caDSR tool for annotating UML Domain Model classes, attributes, associations and enumerations with NCI controlled terminology.
<a href="#">UML Model Browser</a>	A caDSR tool for viewing the caDSR 11179 content that have been recorded in caDSR from an XMI representation of a UML Domain Model. The UML Model Browser provides a view of this content using the model owners names and definitions, and as a collection of UML classes and attributes.
<a href="#">UML Model Loader</a>	A tool that parses the classes, attributes, associations and enumerations contained in an XMI file representing a UML Model Domain Model. The loader transform and records these UML elements into 11179 items in caDSR. The content can be viewed using the UML Model Browser, the CDE Browser or accessed via any of the caDSR tools and services.
<a href="#">caDSR Sentinel Tool</a>	A tool that allows users to subscribe to receive reports about specific categories/collections of changes in caDSR metadata. For example, creating a subscription to receive a report when any part of a CDE that has been used in a particular UML Model, or Form, is changed/edited. A formatted report is sent to an email account or to a server containing the before and after changes to caDSR metadata and related items.
<a href="#">caXchange Software Development Kit</a>	Generates a set of web and grid service artifacts from a semantically annotated XMI file, including an XML Schema representing the service data, a WSDL, and the deployable client/server files.
<a href="#">Common Logging Module (CLM)</a>	A set of tools that provides a flexible and comprehensive solution for auditing and logging. This tool is used with UML models represented as XML to provide object oriented event logging and log query/view capabilities.
<a href="#">CGMM (CSM to grid migration module)</a>	
<a href="#">cgMDR</a>	A 11179 registry and MS Excel and EA add-ins. The cgMDR includes UIs for creating instances of 11179 registry objects, MS Excel and EA for accessing cgMDR, caDSR, EVS or local terminology concepts to create and use 11179 content from within MS Excel and EA.
<a href="#">XMI Handler Archive</a>	A tool for reading and writing to an XMI file. This tool is used by SDK, caAdapter, UML Loader, SIW and Bulk Loader.
<a href="#">Admin Tool Archive</a>	A tool for creating caDSR content, lookup tables and user accounts.
<a href="#">Bulk Loader Archive</a>	A tool that reuses the UML Loader to transform and record caDSR content from annotated MS Excel spreadsheets.
<a href="#">caDSR Services</a>	A set of web, html, java, perl and grid services for accessing caDSR content.
<a href="#">Freestyle API</a>	A UI and API that supports query and weighting of results based on 11179 metamodel, across multiple caDSR objects and attributes simultaneously, regardless of type. For example, using an identifier, find the caDSR item (regardless of whether it's a CDE, VD, OC, etc), weight/order search results based on the type of content. This api is used by the SIW.
<a href="#">Object Cart</a>	An interface that provides a 'shopping cart' feature for use with caDSR or other products that have a need to store objects for later processing, for example, in the CDE Browser and Form Builder to pass items from CDE Browser search results to Form Builder for use in describing forms.
<a href="#">Domain Class Browser</a>	An SDK generated interface for browsing caDSR content via the caDSR and UML Project Domain Models. It supports the same QBE query style as the SDK generated services and returns results in either HTML or XML format.