

Knowledge Repository Use Cases

Contents of this Page

- [Introduction](#)
 - [Background](#)
 - [Scope](#)
 - [Project Scope](#)
 - [Scope Statements](#)
 - [S1: ISO 11179 ed. 3 Registry Support](#)
 - [S2: Model Repository Support](#)
 - [S3: ISO 21090 Support](#)
 - [S4: Semantic Transformation Support](#)
 - [S5: Distributed, Federated Repositories](#)
 - [S6: Semantic Service Oriented Architecture](#)
 - [S7: Semantic Integration Tooling](#)
 - [S8: Semantic Infrastructure Interoperability](#)
 - [Inter-Initiative Touch-Points](#)
 - [Actors and Roles](#)
 - [Related Documentation](#)
- [Manage Model - Record/Create Model Elements](#)
 - [Record New Assertion](#)
 - [Record Rules](#)
 - [Record queries](#)

Introduction

The NCI-supported caBIG program is facing a major expansion of its mission and of the community of stakeholders that it must engage. The success of the program to date has been widely recognized by numerous thought leaders, technology innovators, and program managers in the fields of biomedical computing and health information technology. These leaders have been influenced by both the technological sophistication and the open community processes of the caBIG program, and are seeking to adapt caBIG to meet the much broader agenda of health care information management, sharing, and integration across institutions, communities of practice, and patients.

As a means towards engaging this larger community and contributing to the creation of solutions for its needs, the BIG Health Consortium was formed, a non-governmental organization that is affiliated with caBIG, but which has a complementary agenda that is focused on enabling the integration of research, care delivery, and patient-driven decision making in practical ways. As a consequence, the institutional stakeholders associated with BIG Health Consortium extend well beyond the cancer centers, universities, and cooperative groups that make up the core NCI-funded constituencies that have participated in caBIG since its original incarnation. The extended community thus goes beyond what is considered caBIG-proper today, and includes groups that are not yet connected to caGrid.

Background

The primary goal for this Architecture, Development, and Deployment of a Knowledge Repository and Service is to address the needs of this extended community for a scalable, decentralized infrastructure for managing and disseminating operational metadata and information models, and their associated semantic constructs. The vision for the project is to re-imagine the caBIG technology environment as a more open and more readily extensible framework, one that can grow with less dependency on the centralized processes and systems that are manifest in the first generation of caBIG technology. In particular, the role of the central metadata registry, the caDSR, must be redefined as a federation of metadata registries that can be instantiated and plugged into the caBIG grid or an extended community "cloud" by any qualified entity.

The caDSR has a suite of tools and APIs that support workflows for metadata development, browsing and retrieval. In addition, the caDSR has been adapted to support the UML model-driven development paradigm adopted by caBIG. UML-defined information models such as those from the BRIDG project, caArray, caTissue, and others are each registered in the caDSR through conversion of the model elements into ISO11179 metadata constructs. This functionality, and the workflows that it supports, has evolved over an 8-year period and is now quite mature. It satisfies the current requirements for semantic representation in the current caBIG developer and user community, but it is ill-suited to serve the new requirements for decentralization and indefinite scalability in the broader health care community. The goal of this program is therefore to harvest and recycle the best elements of the first generation of caBIG metadata infrastructure, and to then incorporate those elements into a redesigned and modernized technology stack that is engineered from the start to support a federated deployment topology with far less centralized administration.

Scope



DRAFT

Project Scope

The Knowledge Repository (KR) project is charged with supporting the activities defined in [Initiative 1](#) of the caBIG Semantic Infrastructure effort. Therefore, [the document on the Vocabulary Knowledge Center Wiki that describes Initiative 1](#) is the starting point for defining the high-level scope of this project. The NCI caBIG Semantic Infrastructure and Operations (SIO) group's representatives (Denise Warzel & Dave Hau), referred to here as NCI representatives, will provide direction and ensure that we interpret Initiative 1 in a way that aligns with the vision and mission of SIO, caBIG, and NCI.

The scope statements below represent our **current** understanding of the scope of this project. However, as we gain further understanding of the needs of the [NCI stakeholder community](#), these scope statements will evolve. Therefore, these statements serve to guide our near- to mid-term activities, but do not constrain the ultimate scope of the project.

The scope for **Release 1**, which will be delivered in March of 2011, will be specified in the [Release 1 Scope document](#). That document is derived from the prioritization of lower-level requirements that are specified in the [KR Software Requirements Specification document](#). Furthermore, there are multiple touch-points among initiatives. The [#Inter-Initiative Touch-Points](#) section below describes those and how they affect the scope of this project.

Scope Statements

- [S1: ISO 11179 ed. 3 Registry Support](#)
- [S2: UML Model Repository Support](#)
- [S3: ISO 21090 Support](#)
- [S4: Semantic Transformation Support](#)
- [S5: Distributed, Federated Repositories](#)
- [S6: Service Oriented Architecture](#)
- [S7: Semantic Integration Tooling](#)
- [S8: Semantic Infrastructure Interoperability](#)

S1: ISO 11179 ed. 3 Registry Support

Description:

Provide design and reference implementation of a ISO 11179 ed. 3 compliant metadata repository.

Source:

Initiative 1 explicitly calls for "a production realization of the PIM of the distributed ISO 11179 Ed3 repository..."

Justification:

According to the standard, the purposes of ISO/IEC 11179 are:

- Standard description of data
- Common understanding of data across organizational elements and between organizations
- Re-use and standardization of data over time, space, and applications
- Harmonization and standardization of data within an organization and across organizations
- Management of the components of data
- Re-use of the components of data

To the extent that ISO 11179 achieves these purposes, it provides support for data integration. This aligns with caBIG goals of supporting translational research, which benefits from data integration.

S2: Model Repository Support

Description:

Provide the design and reference implementation of Model Repository, initially to include representations of UML Models, but also to allow recording /registering of other types of end user models that represent the information model for a given application.

Source:

Initiative 1 calls for "...model repositories and tooling including storage and sharing of DAMs and of 'lower-level' models derived from DAMs." It describes the need to use UML-based information models such as BRIDG and LS-DAM to provide a "shared semantic view." It describes the need to enable "model repositories to be decomposed into data description..." We interpret this to mean that there should be some alignment of UML information models to ISO 11179 metadata elements, similar to what is currently in use within caBIG. The purposes for this alignment include 1) encouraging reuse of data elements; 2) enabling description of semantic relationships among UML models. However, alternatives should be considered that may not include ISO 11179, but do support creation of a shared semantic view.

Justification:

BRIDG and LS-DAM represent the consensus, shared semantic views of the clinical research and care and life sciences domains. By providing the capability to describe the semantic relationships among these models and other UML-based information models, the repository would enable view-based data integration, which can support query re-writing or construction of data warehouses.

Additional Notes

One of the goals of this effort is to reduce barriers to sharing metadata. Therefore, the repository should not be restricted to recording only UML or 11179 metadata elements. Instead, it should allow users to record metadata in any form (e.g. spreadsheets, XSDs, etc.). The ISO 11179 ed. 3 model allows for this kind of unconstrained recording of metadata. So, ultimately, we may be able to use that feature of 11179 in order to achieve the goal of lowering barriers.

S3: ISO 21090 Support

Description:

Extend the 11179 Value Domain to support representation of ISO 21090 Healthcare datatypes and accommodate UML models that use these datatypes.

Source:

Initiative 1 does not explicitly call for ISO 21090 support. However, it might be inferred from a call for support of DAMs such as BRIDG which use these datatypes. The Statement of Work for the Knowledge Repository project specifically calls for "Extension of ISO 11179 Value Domain Datatype to include representation of ISO 21090 Healthcare data types" and "support for representation of complex ISO 21090 Healthcare data types" by the services that this project produces.

Justification:

According to the ISO 21090 standard, its purpose is to:

- provide set of data type definitions for representing and exchanging basic concepts that are commonly encountered in healthcare environments in support of information exchange in the healthcare environment,
- specify a collection of healthcare related data types suitable for use in a number of health related information environments,
- declare the semantics of these data types using the terminology, notations and data types defined in ISO 11404 rev 2005,
- provide UML definitions of the same data types using the terminology, notation and types defined in Unified Modeling Language (UML) version 2.0,
- define an XML (Extensible Markup Language) based representation of the data types suitable for use when exchanging information between information processing entities.

To the extent that use of these datatypes will support interoperability, this item aligns with the caBIG goal to "Connect scientists and practitioners through a shareable and interoperable infrastructure."

Additional Notes

The system should provide support for selecting and describing specific 21090 types, and allow localization of 21090 data types (constraining /expanding).

S4: Semantic Transformation Support

Description:

The system should allow metadata that has been stored in the repository to be queried and manipulated from the perspective of multiple views. Therefore, we will define isomorphic transformation between all supported views. ISO 11179 and UML are the currently identified views.

Source:

Initiative 1 calls for "...transformation capabilities to allow model repositories to be decomposed into data descriptions that can be reused through the existing infrastructure to support deployment of these semantics in practical end user solutions via software engineering techniques such as forms development and forms generation."

Here, the main requirement seems to be, enabling use of the existing infrastructure, which is oriented toward a ISO 11179 view of metadata.

The RFP expands the notion of transformation to include higher-level services such as comparison. This is captured here in a separate scope item.

Justification:

There are existing tools (e.g. form builders) that rely on a ISO 11179 view of metadata. This project should support them.

S5: Distributed, Federated Repositories

Description:

The architecture of the 11179 and UML repositories should enable physical distributed and logical federation. The architecture should support distributed linking of content and federated workflows (e.g. versioning, curation, query) over that content.

Source:

The sub title of Initiative 1 is "Distributed, federated metadata repositories and model repositories and operations." It indicates that the "...architecture for the distributed metadata repository (MDR) will be decentralized in nature, allowing multiple peer repositories to be present at the same time, for sharing of data elements."

Justification:

According to the [SI Conops Initiatives overview](#), "The need for all semantic metadata to be formally recorded in a single central repository would limit or preclude application of the semantic infrastructure to very large, diverse communities such national health care. Distributed, federated metadata resources will clearly be required."

S6: Semantic Service Oriented Architecture**Description:**

The architecture of the metadata and model repositories will be service-oriented, according to the principles defined [here](#). Functionality will be exposed through semantically annotated service interfaces.

Source:

The SI Conops calls for the use of the Services Aware Interoperability Framework (SAIF), which prescribes a SOA-based approach.

Justification:

NCI is using SOA as its [strategy](#) to "ensure working interoperability between differing systems that need to access or exchange specific classes of information and/or coordinate cross-application behaviors."

S7: Semantic Integration Tooling**Description:**

This project will produce tooling that utilizes the distributed, federated metadata and model repository services to support an updated and revised semantic integration workflow (i.e. the ECCF) that enables decentralized and localized additions of models and data elements.

Source:

Initiative 1 calls for "Development of a set of user applications or services as needed for creation, management, search and retrieval and of metadata." And, the [SI Conop Mission statement](#) calls for:

- *Employ the Enterprise Compliance and Conformance Framework to represent frameworks and models in an implementation independent manner;*
- *Build and adapt tools and interfaces for generation, curation, storage and use of semantic information, and for convenient lookup, retrieval and transformation of this information by both end-users and applications;*

Justification:

The [semantic infrastructure requirements](#) have both run-time and design-time aspects. The primary design-time requirements involve application of the ECCF to build systems that achieve Working Interoperability.

S8: Semantic Infrastructure Interoperability**Description:**

The NCI's approach to metadata registries should support some level of interoperability with registries of other agencies and external groups.

Source:

<Source>

Justification:

<justification>

Inter-Initiative Touch-Points

<working>

Actors and Roles

An Actor models a type of role played by an entity that interacts with the subject (e.g., by exchanging signals and data), but which is external to the subject. Actors may represent roles played by human users, external hardware, or other subjects. Note that an actor does not necessarily represent a specific physical entity but merely a particular facet (i.e., "role") of some entity that is relevant to the specification of its associated use cases. Thus, a single physical instance may play the role of several different actors and, conversely, a given actor may be played by multiple different instances.

While UML 2 does not permit associations between Actors, this constraint is often violated in practice since the generalization/specialization relationship between actors is useful in modeling overlapping behaviours between actors. The actors below are represented as having a hierarchical relationship for ease of understanding; however, these relationships can easily be removed.

- **Cancer Researcher:** plans and performs activities related to discovery of new knowledge, drugs, and treatments in the field of oncology
 - **Clinical Researcher:** works directly with patients and/or patient data while performing cancer research
 - **Basic Science Researcher:** works with scientifically generated data while performing cancer research
 - **Protocol Designer:** defines the methods used to perform cancer research
- **Information Technologist:** designs, develops, and manages the software and hardware necessary to perform cancer research
 - **Business Analyst:** analyzes the business processes and describes the goals and activities of the user community
 - **Information Modeler:** designs, defines, and describes the data that will be captured during cancer research activities
 - **Software Engineer:** implements the software systems that are used to manage and perform cancer research
 - **Systems Architect:** designs the software and hardware systems that are used to manage and perform cancer research
 - **System Administrator:** manages the software and hardware systems that are used to perform cancer research
- **Metadata Specialist:** having a deep understanding of semantics and syntactic, assists Cancer Researchers and Information Technologists on modeling and managing metadata
 - **Forms Author:** constructs data collection forms based on a library of common data elements
 - **Metadata Curator:** works hands-on with Cancer Researchers and Information Technologists to model and manage their metadata
 - **Metadata Systems Specialist:** manages centralized metadata systems and assists with the design of metadata systems
 - **Terminologist:** a metadata expert that manages and maintains the semantic concepts that underlie information models
 - **Compatibility Reviewer:** while the nature of "caBIG Compatibility" may change as need dictates, the role of the Compatibility Reviewer will likely continue to be a Metadata Specialist that reviews a variety of documents/artifacts to determine the level of interoperability that a system meets
- **Patient:** any person who receives medical attention, care, or treatment
 - **Subject:** a Patient that is participating in Cancer Research

Related Documentation

End User	Analysis	Technical	Management
Knowledge Repository Project Page User Manual Release Notes Installation Guide Developer Guide API Document	Requirements Specification Use Cases	Architecture Guide CFSS PSM PIM	Vision and Scope Roadmap Project Plan Work Breakdown Structure Product Backlog Sprint Backlogs Communications Plan Test Plan Risk Matrix

Manage Model - Record/Create Model Elements

This class of use cases deals with the creation of model elements and recording parts of models for discovery and reuse.

Record New Assertion

Use Case Number	—
Brief Description	—
Actor(s)	—
Pre-condition	—
Post condition	—

Steps to take	—
Alternate Flow	—
Priority	—
Associated Links	—
Fit criterion/Acceptance Criterion	—

Record Rules

Use Case Number	—
Brief Description	—
Actor(s)	—
Pre-condition	—
Post condition	—
Steps to take	—
Alternate Flow	—
Priority	—
Associated Links	—
Fit criterion/Acceptance Criterion	—

Record queries

Use Case Number	—
Brief Description	—
Actor(s)	—
Pre-condition	—
Post condition	—
Steps to take	—
Alternate Flow	—
Priority	—
Associated Links	—
Fit criterion/Acceptance Criterion	—