Semantic Infrastructure Concept of Operations Background

Contents of this Page

Semantic Infrastructure Concept of Operations Links

- Semantic Infrastructure
- Vision
- Background
- Mission
- Objectives
- Overview
- Initiatives
- Stakeholders
- Alternatives

Organization

The cancer Biomedical Informatics Grid (caBIG®) initiative, overseen by the National Institute of Health's (NIH) National Cancer Institute (NCI) Center for Biomedical Informatics and Information Technology (CBIIT), was conceived to connect the cancer research and clinical community through open, interoperable technology that facilitates data sharing and collaboration, speeding the translation of discoveries from research to clinical care. Since its inception, caBIG® has launched innovative tools, infrastructure and policy resources that enable individuals and organizations in the cancer community and beyond to move further down the path to realizing personalized medicine and integrated care. Over 1,000 individuals from more than 190 organizations are currently participating.

Approach

The semantic model and infrastructure forms a key component of the caBIG® collaborative infrastructure, which enables eliciting, capturing, advertising, and discovering the integration semantics for interoperable components and data that may be leveraged to build applications and ultimately to meet user requirements. The current semantic infrastructure uses the ISO 11179 Ed3 standard as its central component in allowing data elements and models to be annotated with concepts, and curated and registered in a repository allowing lookup and retrieval by both end-users and applications. It enables the automatic integration and transformations of data for sharing and collaboration by provisioning the infrastructure with clear, computable, and unambiguous data descriptors for those who would create software that can use and interpret the data in the service of cancer research.

Issues

There are two issues with this broad approach. Currently, we compel developers to do a lot of work to register semantics because of the current design of caGrid which conflates semantic and syntactic concerns. The semantic and syntactic worlds are explicitly separated under ECCF, but their conflation is in our current caGrid semantics is one of the main causes of the two issues noted here.

1. Creation and curation of the needed data semantics descriptors for seamless integration of data via caGrid and secondary use of data descriptions for prospective studies can be demanding, and may ultimately limit the growth and adoption of this aspect of caBIG®.

That is, achieving economies of scale via the semantic infrastructure is desirable but very labor intensive, especially from the standpoint of defining information elements, capturing the essential semantics, and then advertising these captured semantics for reuse. And of course, these information components must be contextualized within one or more behavioral models that express the means by which information is both exposed and the means by which it is analyzed. Thus, caBIG® semantics covering exchanging research information must be capable of expressing and using computationally tractable representations of the interactions among participants engaged in health care and research. Historically, these have largely been ignored in favor of exposing sharable data in caBIG® through the use of compatibility guidelines that defined a leveled approach to promoting optimal exchanging of research information. The refinement of these compatibility guidelines is still underway, but that effort itself has proved difficult. The reasons for this difficulty illuminate the second issue.

2. The semantics infrastructure that has proved remarkably successful in optimizing data exchange in support of cancer research does not automatically provide a similar optimization in other spheres. Thus, the ECCF defined "Cancer Research Data Exchange" as a sphere unto itself, with its own set of optimization paths, but without necessarily being applicable to other spheres (or what the ECCF calls topics). It makes sense to expose objects that represent research domain concepts via web services, as long as the data is stored that way, and to allow interactions that support their direct query. This model (querying semantically precise object graphs without regard for location or persistence type) is adequate in a high trust environment where data integration and analysis is the primary focus, but when workflow and business rules need to be imposed after the results of a query are returned, the requirements are different. The semantic infrastructure must extend to multiple topics where basic data elements are contextualized in ways specific to each research objective, but the existing infrastructure cannot do it. Efforts to develop generalized enterprise-wide standard information elements from those collected to fit particular research needs are not practical. Most information elements cannot be reused to fit all needs, and attempts to use them as universals results in a stalemate for the enterprise and the curator.

It is critical that the caBIG® semantic model and infrastructure evolve to minimize the effort required from data owners and service providers to advertise comprehensive semantics for their data, services, and syntactic infrastructures, promoting reuse and durability. However, it is also clear that across the broad scope of the NCI, this evolution cannot happen along a single path – it should happen on a topic by topic basis, with optimizations learned from the experience of implementation and design.

Background Summary

In summary, the behavioral semantics, those governing the allowable changes in state of an information object, as well as static semantics that describe the specific information at a particular location in time (which has here-to-fore been the focus of caBIG® semantics) must be elicited, captured, stored, and advertised as appropriate, contextualized by topics. In many cases, these topics represent durable components in their own right, and may thus become durable enterprise services that may embody policies or business rules.