National Cancer Institute

caBIG cancer Biomedical Informatics Grid

HIT Standard Committee Meeting

Public Sector Panel

March 8, 2010

Testimony Submitted by Ken Buetow, Ph.D.
Director, Center for Bioinformatics and Information Technology
National Cancer Institute, National Institutes of Health

Background

21st century biomedicine will connect individuals, organizations, institutions and their concomitant information in a cycle of discovery, development and clinical care (i.e., a "rapid learning health system"), transforming the disconnected sectors into a healthcare system that is personalized, preventive, pre-emptive, and patient-participatory. Such a system supports routine assessment of quality of care, comparative effectiveness, pharmacovigilance, and biosurveillance, and will drive an accelerated and more productive generation of research.

The National Cancer Institute (NCI) of the National Institutes of Health is demonstrating this new biomedical "ecosystem" in the cancer community. The information technology platform underpinning this effort is caBIG® ("ca" for cancer and "BIG" for the Biomedical Informatics Grid). BIG uses current internet and web services technology and deploys existing information standards. While the NCI has implemented BIG to support cancer, like the internet itself, it is a robust, highly generalizable framework. It generates "data liquidity" through a comprehensive collection of semantically-aware software services – that is, open-source software deployable locally or accessible as Software as a Service (SaaS) through cloud computing infrastructure. The services expose a variety of application programming interfaces ranging from simple web services to the semantically-rich caGrid services. These capabilities join the diverse, distributed individual and organizational data, applications, and computational capacity into an electronically-connected, virtual biomedical capability known as "The Cancer Knowledge Cloud."

The modular Cancer Knowledge Cloud services can be "choreographed" or "mashed" to create novel, composite resources. These capabilities can be leveraged to facilitate Meaningful Use of Electronic Health Records by healthcare providers, and can enable those providers to measure and report on Quality.

Over the past 5 years, caBIG® has been creating and deploying these capabilities in partnership with 60+ NCI-designated Cancer Centers and 15+ Community Cancer Centers around the nation to enable them to exchange a wide variety of types of information within their own four walls and externally between and among their institutions. caBIG® teams are working with the advocacy community to utilize this infrastructure to empower consumers to contribute to – and benefit from – the rapid learning health system.

What is your role in supporting meaningful use (MU) and quality reporting?

NCI is charged by the 1971 National Cancer Act (42 USC § 285) with coordinating the Nation's Cancer Program. To this end, NCI plays a key role convening and coordinating members of the academic community, professional societies, and members of the advocacy community. NCI works with formally constituted federal advisory committees, including the National Cancer Advisory Board, the Board of Scientific Advisors, and the Director's Consumer Liaison Group. Through this convening and coordination role, it is in a unique position to assist in creating a coherent definition of meaningful use and quality for the cancer community.

Amplifying this leadership role, The National Cancer Act requires the NCI to:

- "Collect, analyze and disseminate all data useful in the prevention, diagnosis, and treatment of cancer"
- "Take necessary action to ensure that all channels for the dissemination and exchange of scientific knowledge and information are maintained between the National Cancer Institute and the other scientific, medical and biomedical disciplines and organizations nationally and internationally".

To operationalize this mission, the NCI also works in collaboration with the diverse biomedical ecosystem in information technology. Through these collaborations NCI:

- *consumes and develops* conformant standards-based specifications to resolve business problems;
- *validates the viability* of specifications through reference implementations; and
- *deploys and hands off* both its specifications and reference implementations to the Commercial and Open-Source vendor communities.

It is, therefore, within NCI's core mission to assist the cancer community in supporting meaningful use and quality reporting. To that end it is working with the cancer provider, advocacy, and research communities to unify definitions, policy, and standards for meaningful use and quality. In partnership with these communities, NCI is creating reference implementations of these specifications.

What resources, experience, expertise and innovative solutions do you have that could support both the public and private sectors?

The resources, experience, expertise, and solutions available to support public and private sectors as a result of the above include:

- NCI caBIG® program infrastructure
- NCI Enterprise Services
- Center for Biomedical Informatics and Information Technology staff
- Experience in linking NCI-designated Cancer Centers
- caBIG® Knowledge Centers
- Experience in large disparate communities
- Experience in linking divergent data types and sources

NCI caBIG® program infrastructure

As described in the Background section, NCI has diverse connections to the Cancer Community through formal federal advisory committees, communities to which it provides direct support, and by interactions with professional and advocacy organizations. NCI coordinates its biomedical informatics ecosystem efforts through the caBIG® program infrastructure. The program infrastructure provides 10 Workspaces and 18 Special Interest Groups through which the Cancer Community can coordinate its efforts and come to consensus on important topics. Fundamental to this approach is the concept of a virtual community. The caBIG® community convenes via open, web-based teleconferences and coordinates its activities through web sites, listserv's and wiki's (http://caBIG.nci.nih.gov). A weekly schedule of all caBIG® activities is distributed and any interested party is invited to participate. A rich collection of support material ranging from documentation to on-line training and tutorials are shared through these resources. More than 2300 individuals from 740+ institutions actively participate in caBIG®. These participants represent the broad biomedical ecosystem of researchers, providers, consumers, academic, government, and industry (IT, pharmaceutical, etc.)

Additionally, the caBIG® program supports Knowledge Centers and Support Service providers. The Knowledge Centers (http://cabig.nci.nih.gov/esn/knowledge_centers) provide community support for caBIG® resources such as vocabularies, infrastructure, applications, and policies. The Support Service Providers (https://cabig.nci.nih.gov/esn/service_providers) are licensed to provide fee-for-service support to the community for utilizing the caBIG® framework, providing Service Level Agreement contracted level of support. More than 15 commercial entities, ranging from large international systems integration companies to small businesses are part of the Support Service Provider network.

NCI Enterprise Services

caBIG® is implemented through a semantically-aware Service Oriented Architecture (sSOA). The sSOA supports the integration of the divergent data types (e.g. outcomes, clinical encounters, biospecimen annotations, medical images, molecular biology information, etc.) that are collected by the large number of independent organizations widely distributed across the United States and beyond. The sSOA enables the coordination of functionality between the various information systems that reside within those organizations to facilitate collaborative data processing and work flow execution. These services represent modular units that can be assembled to support meaningful use and quality reporting.

NCI is developing the catalogue of Business Capability and associated supporting services necessary to meet the informational and functional integration requirements of its diverse oncology research and care community -- including meaningful use and quality reporting. These services can be implemented in a largely standalone fashion to allow for the rapid creation of new applications through the marshalling of services. Alternatively, they can also be integrated with existing applications to ensure working interoperability between differing systems that need to access or exchange specific classes of information and/or coordinate cross-application behaviors.

NCI Services Inventory. The NCI's current inventory of existing or currently-planned services is shown in Figure 1. For convenience of both planning and deployment, the services are classified into four types loosely based on the CBDI service taxonomy. Infrastructure/Utility services provide essential capabilities (e.g. Security, Semantics, Audit, etc.) required by all other services and applications. Core services (e.g. Person, Protocol Abstraction and Organization) provide support for information components common across multiple Business Capability Services (e.g. Specimen Management or Image Management) and form the bulk of the "business atoms," which most users utilize to interact or interoperate within and across traditional system boundaries.

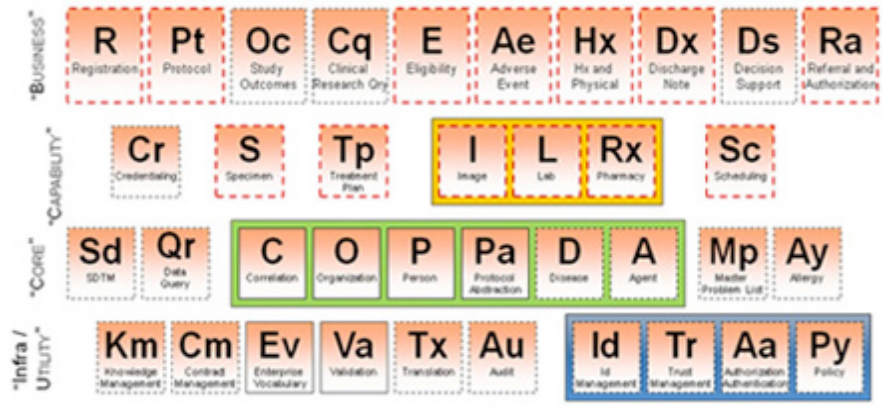

Figure 1: *NCI Enterprise Service Portfolio: Solid outlines indicate services that are in production. Dashed lines indicate services that require specification. Dashed red lines indicate services that are currently being specified under a project within NCI CBIIT.*

Such interactions—particularly when they involve work flow or business processes—utilize a fourth service type called *Process Services* to "orchestrate" or "choreograph" arbitrarily complex, user-specified workflows through combinations of services of the other three service types. NCI services use existing standards whenever possible. In particular, the NCI has adopted the HL7 V3 Reference Information Model (RIM), the BRIDG model (http://www.bridgmodel.org), and the ISO 21090 data type specifications.

Services Implementation Modes. NCI services — both in reference implementation and documented specification format — are available to developers and other interested stakeholders. In particular, detailed specification documents including conceptual, logical and implementable models are available for those that wish to build software that can interoperate with other systems that support NCI services. As is the case for all NCI products, reference implementations of all NCI sSOA services are provided free-of-charge under a non-viral, Open Source license. Finally, the NCI hosts and maintains operating instances of many of these services that can be utilized (with appropriate security) to construct applications requiring access to this data. For maximum ease-of-use, most NCI constructed services surface APIs using multiple technology implementations -- typically an Enterprise Java Bean (EJB) API, a Web Services API, and an API that utilizes caGrid -- the NCI's semantically-aware platform infrastructure. In addition, REST APIs to enable simpler access are available for many of these services.

Service Software Development Kit. A Software Development Kit (SDK) has been created that supports the development of caGrid services -- the caCORE SDK. This SDK generates code to create a "caCORE-like" software system. The code generated is combined with controlled vocabularies and registered metadata to create a "semantically integrated" system in which all exposed API elements have runtime-accessible metadata that define the meaning of the elements using controlled terminology. The caCORE SDK has two modules:

- **Code Generation**, which accepts a UML model as input and produces java classes and other artifacts such as a standard XML Schema (xsd) that can be used by standard client APIs
- **Middleware**, that provides the actual infrastructure including the server and standard client APIs to provide access to the underlying data system thru n-tier architecture

The caCORE SDK is available at: https://cabig.nci.nih.gov/tools/caCORE_SDK

Services-Aware Interoperability Framework (SAIF). In order to scale working interoperability in this technologically diverse environment, the NCI has adopted the HL7 Services-Aware Interoperability Framework (SAIF) as an essential component of its strategy to develop services and other software within an open, distributed Enterprise Architecture Specification. SAIF defines a critical sub-framework — the Enterprise Conformance and Compliance Framework (ECCF) — that enables services to be specified from both an operational and interoperability perspective. The decision to adopt the ECCF was driven by three characteristics of the framework:

- Layered specifications, which allow for binding to multiple technology platforms
- Formal mechanisms for defining site-specific localizations (such as vocabularies or workflows)
- A formal mechanism for embedding, within specification artifacts, Boolean conformance statements that can be used to create testable conformance assertions.

The layered structure of the ECCF enables parties wishing to interoperate via ECCF-specified services to determine with a reasonable degree of certainty the level of difficulty involved in a particular interoperability instance. For example, services/systems that share both technology and payload specification (i.e. they are interoperable at the Platform Specific or implementable level of an ECCF specification) can interoperate more easily and comprehensively than two services/systems whose commonality is only specified at the Platform-Independent (PIM/logical) or Computationally-Independent (CIM/conceptual) level.

In such situations, as mentioned above, the explicit and multi-dimensional nature of ECCF-based specifications enable engineers to quickly and efficiently identify and understand the nature of the problems they must resolve to achieve computable semantic interoperability. Full details of the NCI's Enterprise Services—including a list and description of the services, status of their development and specification documents — is publically available at the NCI Wiki:

(https://wiki.nci.nih.gov/display/EAWiki/Candidate+NCI+Enterprise+Services).

The NCI service development is a community based effort. Individuals and organizations that wish to participate in development of these specifications can contact NCI Application Support at ncicb@pop.nci.nih.gov or by phone at 301-451-4384 or toll free at 888-478-4423 from 8:00AM to 8:00PM Eastern Time.

The NCI is creating tooling to support the generation of SAIF-compliant artifacts, i.e. the artifacts that populate the ECCF Specification Stacks. NCI Center for Bioinformatics and Information Technology (CBIIT) views the development of these tools as a critical success factor for our mission and, as a result, is working on an aggressive timetable to develop these tools. Like the rest of the caBIG® infrastructure, these tools are open-source and will be widely shared with all interested organizations.

Toward Meaningful Use: The NCI Patient Outcomes Data Service and the eHealth for All Partnership

The NCI Patient Outcomes Data Service (PODS) is based on the terminology and standards of caBIG® and the capabilities of the NCI sSOA described above. A first generation prototype has been released, and a production version will be launched in April 2010. PODS enables physicians to collect and share information on the cancer diagnostic, treatment, and clinical outcome of individual patients, as well as the outcomes of all of their patients in the aggregate.

In addition to well-defined APIs that permit electronic transactions, a user application that utilizes the PODS service enables data entry by a provider as well as retrieval and update of a patient's outcome information. PODS therefore provides an interface to private practice physicians as well as hospital-based physicians.

By utilizing the Patient Outcomes Data Service, oncologists will be able to:

- Electronically transfer information from the Electronic Health Record about a patient's diagnosis, treatment interventions and clinical outcome;

- Electronically transfer such a record to the patient within 48 hours of a clinical encounter; and
- Electronically compare contemplated treatments with those that are working most effectively for patients within their own clinical practice, across their entire institution, and eventually across multiple institutions nationwide.

The open-source PODS and its related specification serves as a reference implementation “common front door” for vendor solutions supporting transfer of diagnostic, treatment, and clinical outcome information.

The ”common front door” concept is being demonstrated in the **eHealth For All** (eHFA) Partnership. In this partnership, the NCI Center for Biomedical Informatics and Information Technology (CBIIT), **SAIC Health Solutions Business Unit** (HSBU) and **Microsoft’s Health Solutions Group** (HSG) are collaborating to develop the technical and solution framework necessary to demonstrate the ability of clinicians and cancer survivors to engage in the collection and sharing of provider-reported, consumer-controlled outcome data. These data will be utilized by all members of the cancer community to provide better care, more informed decision-making and more productive research in numerous ways:

- Consumers, motivated by advocacy programs, utilize an eHFA branded portal to access provider-generated and patient-reported health information via Microsoft’s HealthVault.
- Providers, as directed by consumers, are able to utilize PODS to electronically send structured clinical information to a Microsoft-based Amalga data warehouse and a consumer-facing HealthVault resource.
- Providers and consumers gain summary information via easily accessible analytics and visualization tools.
- The aggregated Amalga resource interconnects with the Cancer Knowledge Cloud.

The eHFA project advances us significantly towards the creation of the meaningful use capability of timely consumer electronic access to provider-generated care records.

Describe your roadmap for moving from where you are today to demonstrating meaningful use.

Oncology-extended Electronic Health Record

NCI is collaborating with the American Society of Clinical Oncologists (ASCO) to develop an oncology-extended Electronic Health Record that will facilitate improved care of cancer patients. To date the collaboration has produced the CORE (Clinical Oncology Requirements for the EHR) specification:

- Describing the functions that oncologists want an EHR to perform;
- Providing the structured data elements to be used in oncology EHRs; and
- Defining a common set of interoperability standards that will allow oncology-specific data to be shared from one EHR to another.

Critical to this activity is the definition of meaningful use and quality and the translation of these definitions into specifications and implementations.

In partnership with the cancer and vendor community, the NCI is creating these detailed specifications for a comprehensive set of capabilities necessary to support cancer care delivery. These specifications leverage the large volume of existing specifications related to electronic health. As part of the specification efforts NCI will generate componentized reference implementations to determine the robustness of the specifications and test harnesses for determining conformance and compliance to the specification. These open source modular units can be leveraged by the vendor community to add oncology-related functionality to their EHRs. They can be used as “common front door” services to facilitate interoperability. The NCI’s National Community Cancer Centers will serve as a “natural laboratory” for this oncology-extended Electronic Health Record, since approximately 85% of the cancer patients in this country are seen in community settings. The Community Cancer Centers are partnering with vendors to assist in the addition of these functional modules into their products to address their unique needs.

The NCI Patient Outcomes Data Service is an early member of this comprehensive collection of services that NCI is developing for healthcare providers and researchers.

In the next 12 months, NCI will:

- Release a production version of the Patient Outcome Data Service (PODS)
- Demonstrate the electronic transfer clinical information to consumer controlled resources
- Demonstrate electronic transfer of patient information securely into the Service from Community Cancer Centers, and enable reporting back to the physicians on their aggregated data
- Advance the development of the oncology-extended EHR services
- Initiate additional Services for collection, aggregation, analysis and dissemination of patient information to support Meaningful Use, clinical decision-making, Comparative Effectiveness Research, and quality reporting.