# Deploying the cancer Biomedical Informatics Grid (caBIG<sup>TM</sup>): Taking More Than a Village to Build a Network

## Background

The Cancer Biomedical Informatics Grid<sup>™</sup> (caBIG<sup>™</sup>), an open-source, cancer informatics community initiative, is engaged in the process of implementing technology, infrastructure and policies that will connect the people, institutions, and data in the cancer community. These connections are expected to accelerate the discovery of answers to complex questions of cancer biology and medicine and, thereby, improve patient outcomes. To launch such an initiative across almost 100 institutions, both public and private, requires coordination, and leveraging experts with a variety of backgrounds: from scientists to technologists to lawyers to healthcare managers. This poster will highlight a variety of tools, and products the National Cancer Institute (NCI) funded program has developed to facilitate the deployment of caBIG<sup>™</sup>. Part of this ambitious program has been the government's funding of key deployment leads at the cancer centers to facilitate this deployment. This poster will identify the necessary steps in planning and implementing caBIG<sup>™</sup> technologies, practices, and infrastructure to access and share data. It will also address the need for community experts and resources to aid with the deployment and will address the unique caBIG<sup>™</sup> Enterprise Support Network, built from the cancer informatics community, that support the deployment at cancer centers across the United States. Lessons learned in this program of particular relevance to developing the Public Health Information Network (PHIN) will be highlighted.

## caBIG<sup>™</sup> Vision and Goals

**caBIG™ Vision:** A virtual network of interconnected data, individuals, and organizations whose goal is to redefine how research is conducted, care is provided, and patients/participants interact with the biomedical research enterprise.

Although caBIG<sup>™</sup> is focused around connecting a specific stakeholder community, the overarching goals and principles of the project are widely applicable beyond cancer research:

- > Connecting scientists and practitioners through a shareable and interoperable infrastructure
- > Developing standard rules and a common language to more easily share information
- > Building or adapting tools for collecting analyzing, integrating and disseminating information associated with cancer research and care.

## Self Assessment to Deployment

Assessment				
Aggegginein				
		A		
Certer caBIG" are treated caBIG" are treated self-Assessment Answers for Washington University St. Louis Steman Cancer Certer				
-				
	Question	Range	Answer	
1	Use of caBG** software depends on a means to install and foot a Jana-based software stack with thin web-based clients. Does the institution mantain a certralized data center sufficient to supporting these requirements?	1-5 (1 rome, 5 complete)	4	
2	Are Services Driented Architectures (SCA) provided or supported?	15 (Trane, Swidely)	1Nove	
1	Is there are existing IT Support inflastructure (i.e. help desk) provided?	16 (Trans, Sistergi	3 Parialy	
4	Is then an existing inhastructure for providing baining to end users in applications?	1-5 (1 rune, 5 wel- established)	3 Partially	
6	Is then oparizational coordination diffamorization dispectic vecabularies?	1-5 (1 rune, 5 complete)	11am	
6	Is then organizational coordination of itamonization of specific common data elements?	15 (1 rone, 5 complete)	11km	
7	is them an infrastructure provided for message exchange (e.g. HL7) in the organization?	15 (Trane, 5 widespread)	2	
1	f there is (or ane) nessaging architecture(s) please list them.	Üit .	HL7	
5	Are there formal means for exchange of data between the clinical (hospital) and research data achietes?	15 (1 rome, 5 well- established)	2	
1	Is the computer network bandwidth sufficient to for demanding applications (e.g. imaging or game expression)?	15 (1 rune, 5 excess capacity)	3 Party Sufficient	
11	Are there institutional standards for data and network security?	15 (Trans, 5 wel- established)	4	

Cancer Centers were asked to fill out a 139 question assessment

Center caBIG<sup>™</sup> **Capability Report** 

The sciences reserve.	Antipolety, the report provides a series of decision made source subcrite the function	FORS THE REDUCT DIVISION THREE	enter:
Dectors of this Report			rvational
<ul> <li>Bector 1, Whet Your Deal</li> <li>Bector 2, Infestivolve Res</li> </ul>	niiress		ity Analysis enter in to Self-
	speakly (Techning) and Daffithestipetic		ed to the three
<ul> <li>Sector E Pate to Company</li> </ul>			that the Center
personnel, and commoniant, symmetrizes the Center's suit	ined, peoplical representation of your Denter's countainty in terms of rectinology. Infection Using this was exercing point, each Denter car control, point as access fectors for its in ong oppositions in ansas where call of <sup>th</sup> infectively. In a classification frameway. Areas on	generator. The page below	
Tourisation whereas areas on	ser ti ne tener indicate priema gape o neets in the organization.		functional calorid test and responses
	in fras tructure		
	Technology 04		suggers a calorid y capability racing of
			regh adoption below for beging existing
	Commitment Security		search areas:
	Personnel		ion sider
			vent indicaties that /fanagement tools
Assessment Area	Gefinition	Reavant Gundans	2 at the same time. 2 as an immediate
Pharmacon Andreas	Business, processes and government that ensure its method information and view of the Const	18.1         2.1 <th>o consider an Iged approach. It</th>	o consider an Iged approach. It

Based on their answers to the questionnaire the cancer centers received a document detailing their capabilities

Center caBIG<sup>™</sup> Goals/ **Implementation Plan** Templates

	caBIG		
Cancer Center caBIG™ Goals Outline			
1	Executive Summary of Center caBIG ** Goals		
	<ul> <li>This section should summarize, in 500 words or less, the overall Research Programs at your Cancer Center and how caBiO<sup>244</sup> will advance and align with these program make research goals. Please highlight the various collaborative research benefits expected trom integrating caBiO<sup>144</sup> at the Center. Specifically, be sure to include in this section:</li> <li>A brief summary of the strategic goals and objectives of the Research Programs at the Center (This should come from existing strategic documentation on the Research Programs at the Center)</li> <li>A brief overview of specific areas where caBiO<sup>244</sup> will offer distinct benefits in furthering the Research Programs at the Center</li> <li>A list of potential biom edical research benefits expected from collaborating with researchers at your Center and other Cencer Centers</li> </ul>		
2	Getting Connected with caBJG ** Mission Statement		
	This is a short, one-pensyraph diatement which should dearly diate what the Cancer Center would like to accomplish through the Getting Connected with cat/of <sup>an</sup> program. It should dearly identify the desired larget state, in a manner appropriate for sharing with institutional stateholders and beyond. Specifically, each Center should explicitly state what type of data sharing will occur via the cat/or <sup>an</sup> intrastructure.		
3	Benefits of caBIG ** Implementation		
	This section should describe how caBIO <sup>™</sup> implementation will help achieve the Cancer Center's overall Research Program goals and objectives.		
4	Relevant ca8IG ** Implementation Opportunities		
	This section should identify where the Center will focus implementation efforts. Although the analysis output from the caBiO <sup>™</sup> Self Assessment, namely, the Center Capability Analysis Report, will provide a comprehensive set of options available to the Center around caBiO <sup>™</sup> intrastructure, tools, and it andards, e ach Center should evaluate the neisvancy of these opportunities against the Center's own goals and mission. This section should describe only those caBiO <sup>™</sup> components that the Center plans to pursue.		
6	caBIG 14 Implementation Goals		
	This section should identify at a high level how the Center will implement the following three cablo?" goals at the Center:		
	-		

The cancer centers were then able to set their goals for implementation and deployment of caGrid and other data sharing tools based on their capability reports

## caBIG<sup>™</sup> Pilot Phase

- > 190 participating organizations
- > 300 software components
- > 40+ end-user applications in discovery, clinical trials management, biospecimen management, etc.
- > caGrid providing data transmission network that "connects" everyone
- > 45+ peer-reviewed publications about our using caBIG<sup>™</sup>.

## caBIG<sup>™</sup> Enterprise Phase

- > 43 of 63 NCI-designated Cancer Centers actively participating in caBIG<sup>™</sup> deployment program
- > Growing engagement with broader cancer biomedical research community.

The caBIG<sup>™</sup> initiative, having successfully completed its Pilot, is now entering an Enterprise (deployment) Phase. Our overarching objective is to connect the people, institutions, and data in the cancer community through caBIG<sup>™</sup>. These connections are expected to accelerate the discovery of answers to complex questions of cancer biology and medicine and, thereby, improve patient outcomes.

## Approach to Achieving Interoperability

- > Four different levels of caBIG<sup>™</sup> compatibility (i.e., interoperability): Legacy, Bronze, Silver, and Gold.
- > There are four areas of compatibility: Information Models, Vocabularies, CDEs and Programming/Messaging Interfaces.
- > An application must meet the guidelines in all four areas to be considered "caBIG Compatible"

#### There are five steps in developing a caBIG<sup>™</sup> compatible application:



Analytical Service Invocation

> Query Execution

User Management

CaBIG The caBIG<sup>™</sup> **Pilot Phase** Report: 2003-2007 

Adopting a tool involves installing software applications already created by caBIG<sup>™</sup>, integrating them into the workflow, and connecting to caGrid.

Adapting a tool involves modifying the existing software applications to be caBIG<sup>™</sup> compatible and then connecting to caGrid.

A variety of technical, product, and training programs are being developed that will facilitate the adoption of caBIG<sup>™</sup> technologies. They include key caBIG<sup>™</sup> resources, packaged into "bundles," that have been designed to help support and streamline clinical trials, imaging, tissue banking, and integrative cancer research, and to provide the materials needed to join the secure caBIG<sup>™</sup> data-sharing framework.

## Design Pattern Overview

The following outlines SIX different approaches for adapting a tool to be caBIG<sup>™</sup> compatible. These are conceptual models—foundational strategies—for shaping an adaptation roadmap. This is neither an exhaustive nor mandated list—it simply provides examples that we believe can facilitate adaptation based on experience to

### **Understanding Design Patterns**

The six approaches offered here are based in the concept of design patterns, which Wikipedia defines as "a general reusable solution to a commonly occurring problem in software design."

Design patterns are essentially templates: standardized conceptual solutions that can be applied to many different situations. Based on caBIG<sup>™</sup> experiences, there are six standard ways to approach the software design problems presented by the adapt path identified.

## Booz | Allen | Hamilton

## Two Pathways—Adopt or Adapt

## caBIG<sup>™</sup> Technologies

Compatibility Achieved through caBIG™ Bundles			
Clinical Trials Compatibility Framework	Life Sciences Distribution	Data Sharing and Security Framework	
<ul> <li>&gt; C3PR</li> <li>&gt; PSC</li> <li>&gt; caAERS</li> <li>&gt; caXchange</li> <li>&gt; CTODS</li> <li>&gt; caGrid</li> </ul>	<ul> <li>&gt; CTODS</li> <li>&gt; caArray</li> <li>&gt; caTissue</li> <li>&gt; geWorkbench</li> <li>&gt; caGWAS</li> <li>&gt; NCIA</li> <li>&gt; caGrid</li> </ul>	<ul> <li>&gt; caBIG<sup>™</sup> Policies</li> <li>&gt; Processes and Best Practices</li> <li>&gt; Model Documents</li> <li>&gt; Trust Fabric</li> </ul>	

#### Wrapper

There is a desire to continue using an existing legacy tool, without adopting one from caBIG<sup>™</sup>. In this case, a new caBIG<sup>™</sup> API is generated to allow data exchange between the existing tool and caGrid.

#### **7** Direct Data Access (Interim Solution)

There is a desire or need to retain the legacy database, but also willingness to adopt a caBIG<sup>™</sup> tool for its user interface and to facilitate connection to caGrid.

#### Message Broker

The existing tool already uses standardized messages, such as HL7, to facilitate dynamic information exchange between diverse tools.







#### **1** Data Warehouse

There is a desire or need to retain the legacy database and user interface, but there is a willingness to use a caBIG<sup>™</sup> tool to connect to caGrid.

#### **5** Clone and Own

There is a desire or need to retain the legacy database and user interface, and willingness to use a caBIG<sup>™</sup> tool API to connect to caGrid.

#### **6** Generate an API

There is a desire to retain the legacy database, and no existing caBIG<sup>™</sup> API is appropriate to map from the legacy database to caGrid. Instead, caBIG<sup>™</sup> metadata descriptions and CDEs are used to generate a new API to rest on the legacy tool.



caGrid
caBIG™ API
DB
UI
Existing Tool











## caBIG<sup>™</sup> Enterprise Support Network

As caBIG<sup>™</sup> moves into the next phase of development and deployment, the new caBIG<sup>™</sup> Enterprise Support Network (ESN) will help 21st century biomedical researchers integrate caBIG<sup>™</sup> tools and infrastructure at their institutions. The ESN represents a major step forward, by authorizing a variety of external organizations, both academic and commercial, to deliver support services in their particular areas of expertise. The ESN will provide wide-ranging support for caBIG<sup>™</sup> standards, applications, and infrastructure to individuals, including end users, administrators and IT staff; organizations and institutions of the current caBIG<sup>™</sup> community; the broader cancer research enterprise, and eventually others in the biomedical community from other domains who can benefit from interoperability. Collectively, the programs of the ESN can help individuals and organizations determine which technologies to adopt or adapt and how to integrate them into their research enterprise.

#### At a Glance (as of June 2008)

Functions	Knowledge Centers	Service Providers	Program Offices
Contribute to community Web site in their domain			
Contribute to com- munity Web site in their domain			
Training and education			
Provide dev support and enhancements to caBIG™ tools			
Provide help desk and phone support			
Provide server hosting and maintenance			
Provide on-site support for deployment			
Conduct Internal assessment of caBIG™ deployment			
Coordinate inter-departmental and/or inter-org usage			

## Lessons Learned/Implications for Public Health Grid Deployment

- > We have openly and transparently engaged and collaborated with a diverse stakeholder community in order to create an interoperable cancer research data information network to promote efficient electronic information sharing in order to protect health and save lives
- > As an example, caBIG<sup>™</sup> has shown how information systems can support typical public health activities
- > Facilitate collaborative research through greater access to public health research data made available through caBIG<sup>™</sup> infrastructure
- > Increase the research potential from data sets, by facilitating the flow of data between multiple tools on the "bench to bedside" path
- > Install free customizable caBIG<sup>TM</sup> software applications to support the conduct of basic and clinical cancer research
- > The increased data sharing made possible by Getting Connected with caBIG<sup>TM</sup> increases the effectiveness and efficiency of cancer research, helping individual scientists, the cancer research community, and -ultimately – the public's health.

## Looking Ahead: Facing Challenges Together

- > The vision and need for caBIG<sup>TM</sup> have been established, and tremendous achievements made. Despite this, challenges remain:
- Educating the broader community about the available tools and their suitability for a particular research or clinical need
- Creating methods and a useful workflow for data sharing between basic science and clinical research
- Reconciling data models and terminology across traditionally separate areas of science and clinical practice
- Addressing issues of data sharing, including ensuring data security, patient privacy, and intellectual capita
- Marshalling the diverse resources and standards developed in a variety of settings, and organizing them to address the common problem of cancer research and care
- Coordinating and correctly sequencing multiple interdependent software and standards development projects to meet broader community needs.
- > As with any innovative and comprehensive initiative, the challenges of caBIG<sup>™</sup> provide the opportunity for collaborative solutions that, in time, will reshape the overall cancer research and treatment paradigm.