

**AN ASSESSMENT OF THE IMPACT OF THE  
NCI CANCER BIOMEDICAL INFORMATICS GRID (caBIG<sup>®</sup>)**

**Report of the Board of Scientific Advisors  
Ad Hoc Working Group**

**March 2011**

**NATIONAL INSTITUTES OF HEALTH  
National Cancer Institute  
Board of Scientific Advisors**

Ad hoc Working Group on the NCI  
Cancer Biomedical Informatics Grid (caBIG<sup>®</sup>) Program

**Andrea Califano, Ph.D. (Chair)**

Professor of Systems Biology  
Departments of Biochemistry and Molecular  
Biophysics and Biomedical Informatics  
Founding Director and Chair, Columbia  
Initiative in Systems Biology  
Director, Sulzberger Columbia Genome  
Center  
Associate Director for Bioinformatics,  
Herbert Irving Comprehensive Cancer  
Center  
Columbia University Medical Center  
New York, NY

**Arul Chinnaiyan, M.D., Ph.D.**

Professor  
Department of Pathology and Urology  
Director, Pathology Research Informatics  
Director, Cancer Bioinformatics  
Michigan Center for Translational Pathology  
Ann Arbor, MI

**Geoffrey M. Duyk, M.D., Ph.D.**

Partner and Managing Director  
TPG Biotech  
San Francisco, CA

**Sanjiv S. Gambhir, M.D., Ph.D.**

Virginia and D. K. Ludwig Professor  
Director, Molecular Imaging Program  
Director, Canary Center at Stanford for  
Cancer Early Detection  
Head, Nuclear Medicine  
Professor, Departments of Radiology &  
Bioengineering Stanford University  
Stanford, CA

**Tim Hubbard, Ph.D.**

Head of Informatics  
Wellcome Trust Sanger Institute  
Wellcome Trust Genome Campus,  
Hinxton, Cambridgeshire. CB, UK.

**David J. Lipman, M.D.**

Director, National Center for Biotechnology  
Information  
National Library of Medicine, NIH  
Bethesda, MD

**Lincoln D. Stein, M.D., Ph.D.**

Director, Informatics and BioComputing  
Platform  
Ontario Institute for Cancer Research  
Toronto, ON, Canada

**Jean Y Wang, Ph.D.**

Distinguished Professor of Medicine  
University of California, San Diego  
School of Medicine  
Associate Director of Basic Research  
Moores UCSD Cancer Center  
La Jolla, CA

**Executive Secretary**

**Olivia T. Bartlett, Ph.D.**

Chief, Research Programs Review Branch  
Division of Extramural Activities  
National Cancer Institute, NIH  
Bethesda, MD

**Committee Management Officer**

**Ms. Claire L. Harris**

Division of Extramural Activities  
National Cancer Institute, NIH  
Bethesda, MD

## TABLE OF CONTENTS

Executive Summary	i
Report of the BSA Ad Hoc Working Group	1
I. Charge to the Ad Hoc Working Group	1
II. Introduction	1
III. Overview of the caBIG <sup>®</sup> Program	2
IV. Findings	4
Creation and Management of Standards for Data Exchange	5
Impact and Track Record of caBIG <sup>®</sup> Initiatives	6
Life Science/Integrative Cancer Research Tools	6
Clinical Data Management System	9
Infrastructure Tools	11
Community Engagement	13
caBIG <sup>®</sup> Program Administration, Contracts Management and Budget	14
V. Conclusions	17
VI. Recommendations	21
Appendix A: List of WG Meeting and TC dates	
Appendix B: Agenda and List of speakers for WG Meeting on November 2, 2010	
Appendix C: caBIG <sup>®</sup> Connectivity and Usage Chart (provided by caBIG <sup>®</sup> )	
Appendix D: List of Institutions Interviewed	
Appendix E: Diagram of caBIG <sup>®</sup> tools and Workspaces (provided by caBIG <sup>®</sup> )	
Appendix F: List of caBIG <sup>®</sup> Service Providers (provided by caBIG <sup>®</sup> )	
Appendix G: Partial list of caBIG <sup>®</sup> tools in each category (provided by caBIG <sup>®</sup> )	
Appendix H: caBIG <sup>®</sup> Organizational Chart (provided by caBIG <sup>®</sup> )	
Appendix I: caBIG <sup>®</sup> Budget Summary – Fiscal Years 2004 – 2010 (provided by caBIG <sup>®</sup> )	

## EXECUTIVE SUMMARY

Since his appointment in July 2010, the NCI Director has undertaken a review of NCI's largest programs. Over the last seven years, the Cancer Biomedical Informatics Grid (caBIG<sup>®</sup>), supervised by the NCI's Center for Bioinformatics and Information Technology (CBIIT), has been one of NCI's most far-reaching programs, dedicated to designing and developing the next generation of collaborative IT infrastructure for biomedical research. Such an infrastructure would be capable of handling data collection, integration, analysis, and dissemination challenges across the grid of NCI designated cancer centers, collaborating institutions, cooperative groups, and other NCI programs to accelerate the discovery of new approaches for the detection, diagnosis, treatment, and prevention of cancer.

The caBIG<sup>®</sup> budget has grown annually, from approximately \$15 million in fiscal year 2004 to more than \$47 million of appropriated money in fiscal year 2010. An additional \$87 – 100 million from the American Reinvestment and Recovery Act (ARRA) in fiscal years 2009 and 2010 brings the total cost of the caBIG<sup>®</sup> program to at least \$350 million for fiscal years 2004 to 2010. Future plans, including electronic health records (EHR), cloud computing, and other far-reaching activities related to personalized molecular medicine, are likely to continue the trend of escalating expenses. Therefore, a thorough and objective review of this important NCI program is warranted at this time.

To undertake the assessment, the WG requested information from the caBIG<sup>®</sup> leadership on caBIG<sup>®</sup> program activities in four areas: Life Sciences/Integrative Research tools; Clinical Trials Management tools; Infrastructure/Data Sharing tools; and Budget, Program Administration and Contracts Management. The WG conducted an interview-based assessment of the caBIG<sup>®</sup> program, interviewing 59 individuals with a wide variety of caBIG<sup>®</sup>-relevant experiences and perspectives from 46 institutions. The interviews focused on the impact of the caBIG<sup>®</sup> program on the NCI-designated Cancer Centers, the cooperative clinical trials groups, and other NCI research initiatives which caBIG<sup>®</sup> was expected to support, such as The Cancer Genome Atlas project, and industry.

The results of this 4-month assessment have been surprisingly uniform and far less polarized than was originally expected. There was complete agreement that caBIG<sup>®</sup>'s original goals were worthy and remain highly relevant to the future of cancer research in the United States (U.S.). However, there was also strong consensus among those interviewed that caBIG<sup>®</sup> has expanded far beyond those goals to implement an overly complex and ambitious software enterprise of NCI-branded tools, especially in the Clinical Trial Management System (CTMS) space. These have produced limited traction in the cancer community, compete against established commercial vendors, and create financially untenable long-term maintenance and support commitments for the NCI. Furthermore, creating this all-inclusive software enterprise has required the support of a vast management network of external contractors that consumed at least \$60M in overhead costs in the past seven fiscal years and continues to grow.

There appears to be only a few NCI-Designated Cancer Centers that have adopted the full caBIG<sup>®</sup> CTMS solution, while adoption of individual components was relegated to small pilot projects, with little impact on the Centers' mainstream operation. Progress on the caBIG<sup>®</sup> Life Science tools has been somewhat better, with a handful of tools being broadly adopted by several

research lab and large projects. However, the level of impact for most of the tools has not been commensurate with the level of investment. For example, many tools, such as caArray (\$9.3M), have been developed at significant expense and without a clear justification, particularly since a number of similar commercial and open software tools already existed. It is indeed noteworthy and a lesson for the future that the more widely adopted Life Sciences tools have their roots in projects that were already fairly successfully developed by academic research institutions, whereas most of the caBIG<sup>®</sup>-initiated projects have been less successful and, ironically, much more expensive. Similarly, enormous effort was devoted to the development of caGRID (\$9.8M), an environment for grid-based cloud computing, but the WG did not find evidence that it has empowered a new class of tools to “accelerate the discovery of new approaches for the detection, diagnosis, treatment, and prevention of cancer” as envisioned.

The WG’s analysis also revealed problems in the approaches used by the program for implementing the highly valuable vision it had helped define. In particular, the interviews suggest that the strategic goals of the program were determined by technological advances rather than by key, pre-determined scientific and clinical requirements. Thus, caBIG<sup>®</sup> ended up developing powerful and far-reaching technology, such as caGRID, without clear applications to demonstrate what these technologies could and would do for cancer research. While some large projects, such as the I-SPY Breast Cancer study, have been built around caBIG<sup>®</sup> tools, the WG struggled to find projects that could not have been implemented with alternative less expensive or existing technologies and software tools.

Perhaps the greatest impact of the caBIG<sup>®</sup> program on cancer research has been to gather several communities around a virtual table to help create and manage community-driven standards for data exchange and application interoperability. The development of a semantic infrastructure that allows data to be harmonized across cancer centers is widely perceived to be one of the most important contributions of the caBIG<sup>®</sup> program. Importantly, caBIG<sup>®</sup> helped to move the cancer research community beyond messaging systems and limited structured vocabularies and ontologies to push for semantic standards that have achieved significant penetration in the cancer clinical research community. The program has also had impact by supporting the development, maintenance, enhancement, and dissemination of software tools developed by the academic research community.

The WG was surprised to discover that caBIG<sup>®</sup> projects and initiatives have not undergone the usual NCI concept review and approval process, depriving the program of the opportunity to receive valuable guidance in shaping its strategies, approaches and priorities as it grew. Despite the obvious qualifications, technical vision, and integrity of caBIG<sup>®</sup>’s NCI management team, the lack of independent external oversight and the non-peer-review based funding decisions have significantly compromised the ability of the caBIG<sup>®</sup> program to achieve its initial goals.

The WG would like to stress that, going forward, the creation of an infrastructure for data collection, management, analysis, and dissemination remains a critical and only partially addressed problem. It is thus critical that the WG’s findings about the caBIG<sup>®</sup> program’s progress and traction does not diminish NCI’s enthusiasm for and commitment to supporting this critical area of development. Specifically, we recommend that caBIG<sup>®</sup> return to its original mission and premises and that NCI focus separately on informatics tools for clinical and basic research components. The former should become more driven by the requirements of the organizations that run clinical studies. The latter should be better integrated with NCI’s existing

portfolio of programs that support the development of highly innovative analytical tools, which currently lack any but the most basic form of support for community-based software development, maintenance, and dissemination.

The WG also recommends certain immediate actions aimed at reducing expenditures while the program is reorganized, and at creating a critically needed mitigation plan to support the labs and organizations that have become dependent on caBIG<sup>®</sup> tools and that may suffer from the reorganization process.

## **Recommendations:**

### ***Immediate Tactical Recommendations***

1. Institute an immediate moratorium on all ongoing software development projects, both internally within caBIG<sup>®</sup> and through commercial contracts, (such as enhancement and development of tools in the CTMS suite, the caGRID, cloud computing, EHR, and caBIG 2.0) while initiating a mitigation plan to lessen the adverse impact of this moratorium on the cancer research community. Support for maintenance of caARRAY, caTissue, the imaging tools and ongoing multi-site clinical trials dependent on caBIG<sup>®</sup> tools should be exempt from this moratorium.
2. Institute a one-year moratorium on the initiation of all new projects, contracts and subcontracts through caBIG<sup>®</sup> pending their review by the independent oversight committee described in Recommendation 4.
3. Provide a one-year extension of caBIG<sup>®</sup> supported academic efforts for development, dissemination, and maintenance of new and existing community-developed software tools.
4. Establish an independent oversight committee, representing academic, industrial, and government (NCI, NIH) perspectives to review ongoing and planned initiatives for scientific merit and to recommend effective transition options to current users of caBIG<sup>®</sup> tools.
5. Conduct a thorough audit of all aspects of caBIG<sup>®</sup> budget and expenditures to identify unspent funds that can be reprogrammed for use in implementing the WG's other recommendations and for other NCI priorities.

### ***Longer Term Strategic Recommendations***

6. Create a Scientific Advisory Group (SAG) that has an appropriate mix of scientific, technology and informatics expertise to advise NCI on its priorities, future initiatives, business model(s), and resource allocations in the area of biomedical informatics. The SAG should also facilitate abatement of barriers with similar efforts in other NIH Institutes, in the community and abroad. It might be appropriate for a subcommittee of the BSA to do this function.
7. Refocus caBIG<sup>®</sup> on its original mission and discontinue all strategic efforts to develop and maintain its own brand of software tools, either directly or indirectly through commercial contractor efforts.

8. Separate the clinical informatics and bioinformatics components of the caBIG<sup>®</sup> program.
9. Use the usual and established mechanisms for concept clearance through the NCI BSA and peer review of NCI biomedical informatics initiatives in the future.
10. Promote interoperability and data sharing by making them key review criteria for grant and cooperative agreement applications and R&D contracts and by including them as requirements for award.

## REPORT OF THE BSA AD HOC WORKING GROUP

### CHARGE TO THE WORKING GROUP

At its November 2010 meeting, the NCI Board of Scientific Advisors (BSA) of the National Cancer Institute (NCI) voted to create an ad hoc Working Group (WG) to provide an independent review of the Cancer Biomedical Informatics Grid (caBIG<sup>®</sup>) Program. caBIG<sup>®</sup> is an information network that was conceived in 2004 with the goal of enabling all constituencies in the cancer community, i.e. researchers, physicians, and patients, to share data and knowledge. The charge to the WG is:

*... to assess the progress of the caBIG<sup>®</sup> program toward its original goals, as well as to evaluate its accomplishments, challenges, and efforts in community outreach. The goal is to help caBIG<sup>®</sup> achieve even greater traction in the cancer research community and to identify stumbling blocks and areas that will require greater attention in the future development of the program. In particular, the Working Group's evaluation will cover caBIG<sup>®</sup> tools and activities in the following four areas: (a) clinical research tools, (b) analytical discovery tools, (c) research infrastructure, and (d) contract review and administration processes.*

### INTRODUCTION

Since his appointment in July 2010, the NCI Director has undertaken a review of NCI's largest programs. The caBIG<sup>®</sup> program is one of NCI's most far-reaching programs. The data management and informatics activities of caBIG<sup>®</sup> have been a central element in NCI's intramural and extramural programs for at least the last 7 years. The common data vocabularies and formats, clinical informatics tools, discovery tools, and other activities of the caBIG<sup>®</sup> program are expected to broadly impact basic and clinical research activities at NCI-designated cancer centers, community cancer centers, cooperative clinical trials groups, and other research institutions in the US and internationally. caBIG<sup>®</sup> also interacts with other NIH institutes and centers and other agencies in the Federal government.

The caBIG<sup>®</sup> budget has grown annually, from approximately \$15 million in fiscal year 2004 to more than \$47 million of appropriated money in fiscal year 2010. An additional \$87 – 100 million from the American Reinvestment and Recovery Act (ARRA) in fiscal years 2009 and 2010 brings the total cost of the caBIG<sup>®</sup> program to at least \$350 million for fiscal years 2004 to 2010. caBIG<sup>®</sup> initiatives planned for the future, including electronic health records (EHR), cloud computing, and other far-reaching activities related to molecular medicine, are likely to be equally expensive. Therefore, a thorough and objective review of this important NCI program is warranted at this time.

To undertake the assessment, the WG requested information from the caBIG<sup>®</sup> leadership on program activities in four areas: Clinical Trials Management tools; Life Sciences/Integrative Research tools; Infrastructure/Data Sharing tools; and Budget, Program Administration and Contracts Management. The WG also heard presentations from the Director of the NCI Center for Bioinformatics and Information Technology (CBIIT), which manages the caBIG<sup>®</sup> program,



other members of the NCI caBIG<sup>®</sup> senior leadership team, and the two major caBIG<sup>®</sup> contractors, SAIC and Booz-Allen & Hamilton (BAH), in those four areas at an all day meeting on November 2, 2010; Appendix A shows the agenda and list of speakers for that meeting. Following up on the discussions during the meeting on November 2nd, the WG requested additional information to clarify the caBIG<sup>®</sup> budget and usage. The WG held 16 teleconferences between November 22, 2010 and February 25, 2011 to discuss the data and information provided by the caBIG<sup>®</sup> program, interview strategies, interview results, conclusions and recommendations (Appendix B).

The WG conducted an interview-based assessment of the caBIG<sup>®</sup> program, focusing on its impact on NCI-designated Cancer Centers, on the cooperative clinical trials groups, on other NCI research initiatives which caBIG<sup>®</sup> was expected to support, such as The Cancer Genome Atlas (TCGA) project, and on industry. Although the analysis done by the WG between November 2010 and the end of February 2011 was not as comprehensive as a bioinformatics survey in 2008-2009 (*1*), in which 394 individuals in 60 cancer centers were interviewed, the WG members were able to interview a broad cross section of potential caBIG<sup>®</sup> users.

The WG contacted by e-mail almost 75 individuals at 60 institutions, including 43 of the 51 cancer centers listed in the caBIG<sup>®</sup> usage chart presented by the caBIG<sup>®</sup> program leadership during the November 2nd meeting (Appendix C). Over 70% of those contacted responded, and the WG members participated in direct phone interviews of 59 individuals from 46 institutions (Appendix D). These individuals have a wide variety of caBIG<sup>®</sup>-relevant experiences and perspectives, and included:

- Directors of NCI-designated Cancer Centers
- Managers of clinical informatics facilities at NCI-designated Cancer Centers
- Senior researchers and thought leaders in organizations and research labs that either are or are likely to be users and adopters of the caBIG<sup>®</sup> research and infrastructure tools
- Experts in the cancer prevention and health IT arenas
- Leaders of several companies providing commercial software solutions for electronic clinical data capture and management
- Senior investigators at several research-oriented drug development companies.

Both strong supporters and constructive critics of caBIG<sup>®</sup> were interviewed; surprisingly, both made many of the same points. Taken together, the interviews provided a clear and consistent picture of the contributions, track record, impact, challenges and deficiencies of the caBIG<sup>®</sup> program, and an objective basis for the WG's conclusions and recommendations.

## **OVERVIEW OF THE caBIG<sup>®</sup> PROGRAM**

As stated on the caBIG<sup>®</sup> website, the mission of caBIG<sup>®</sup> is to develop a collaborative information network that accelerates the discovery of new approaches for the detection, diagnosis, treatment, and prevention of cancer. caBIG<sup>®</sup> activities are supervised by the NCI Center for Bioinformatics and Information Technology (CBIT). The initiative operates through an open development community. The goals of caBIG<sup>®</sup> are to:

- Connect scientists and practitioners through a shareable and interoperable infrastructure
- Develop standard rules and a common language to share information more easily
- Build or adapt tools for collecting, analyzing, integrating, and disseminating information associated with cancer research and care.

Since its start, caBIG<sup>®</sup> has been committed to the following principles:

- Federated: caBIG<sup>®</sup> software and resources are widely distributed, interlinked, and available to everyone in the cancer research community, but institutions maintain local control over their own resources and data.
- Open development: caBIG<sup>®</sup> tools and infrastructure are being developed through an open, participatory process. caBIG<sup>®</sup> leverages existing resources whenever possible, rather than building new tools in every case.
- Open access: caBIG<sup>®</sup> resources are freely obtainable by the cancer community to ensure broad data-sharing and collaboration.
- Open source: The caBIG<sup>®</sup> source code is available to view, alter, and re-distribute.

The caBIG<sup>®</sup> program had its roots in a project begun as part of the NCI Director's Challenge program in 1999 to develop standards to share analytical tools and to establish principles and rules for data sharing in the cancer research community (2). The original intent was to develop standards for interoperability and analytical tools and to provide a forum for comparison of data related to large scale gene expression data sets. The project was managed by what was then the NCI's Center for Bioinformatics (NCICB).

NCICB gave informational updates to the National Cancer Advisory Board (NCAB) in June 2002 (3) and June 2003 (4) about the evolving program and its goal to build a common architecture and common data standards. caBIG<sup>®</sup> was defined as a collection of interconnected data sources to support cancer research by making data from diverse research disciplines available, integratable and distributable from a variety of nodes in research institutions. A prominent goal was to have all new applications and infrastructure developed in a modular fashion, supporting the creation of complex tools using a "lego block" approach.

A 3-year caBIG<sup>®</sup> pilot project including participants at several institutions was initiated between July 2003 and February of 2004 to "test the feasibility of developing and deploying an integrating biomedical informatics structure" (4, 5, 6). The goals of the pilot project were to demonstrate that (1) a spectrum of Cancer Centers with varying needs and capabilities could be joined in a common grid of communications, shared data, applications, and technologies; (2) Cancer Centers, in collaboration with the NCI, could develop new enabling tools and systems that could support multiple Cancer Centers; (3) Cancer Centers would actively use the grid; and, (4) extensible infrastructure could be expanded and extended to members of the cancer research community. The business model was to "create a front end that will make caBIG<sup>®</sup> attractive for others to invest in and take on responsibility for downstream events" (5) and it was anticipated at the time that the caBIG<sup>®</sup> effort would evolve into a self-sustaining community (7).

As part of the pilot, the caBIG<sup>®</sup> program conducted site visits to 49 of the 61 NCI-designated cancer centers between July 2003 and January 2004 to determine their informatics needs.

Clinical data management tools and databases and staff resources were by far the most common needs articulated by the cancer centers visited.

caBIG<sup>®</sup> leadership presented several informational updates to the BSA (March 2004) and the NCAB (November 2004) during the caBIG<sup>®</sup> pilot to explain the structure of the initiative, including the domain workspaces being formed for Clinical Trials Management Systems, Integrative Cancer Research, Tissue Banks and Pathology Tools, Vocabularies and Data Elements, and Architectural Standards, the 23 Special Interest Groups (SIGs) that were formed to focus on specific topics, and other ongoing activities (5, 6). The caBIG<sup>®</sup> Strategic Plan 2005 (8) included goals and objectives for all of the workspaces for 2005, 2008 and 2010.

caBIG<sup>®</sup> was expanded to a full-scale enterprise initiative managed by CBIIT in 2007. Most of the NCI-designated Cancer Centers received supplemental funds to support caBIG<sup>®</sup> deployment leads (DL) in fiscal year 2007, 2008, and 2009.

As of February 2011, the caBIG<sup>®</sup> program includes the following components:

- Four domain level workspaces, three strategic level workspaces, and two cross-cutting workspaces (Appendix E)
- The caBIG<sup>®</sup>-supported Deployment Program, including deployment leads at 56 NCI-designated cancer centers and a Deployment Advisory Center
- Six caBIG<sup>®</sup>-supported Knowledge Centers to provide demonstrations, training material, answers to frequently asked questions, latest versions of caBIG<sup>®</sup> software and other resources to users of caBIG<sup>®</sup> tools
- Six caBIG<sup>®</sup>-supported In Silico Research Centers of Excellence for investigator-initiated research using data-mining and other in silico methods to investigate etiology, diagnosis, treatment and prevention of cancer
- 19 Support Service Providers, which are licensed commercial entities available to assist users in installing, modifying and using caBIG<sup>®</sup> tools and technologies (Appendix F)

There are currently more than 70 open-source caBIG<sup>®</sup> software tools (Appendix G) for clinical, molecular biologic, imaging, and specimen banking activities. According to data provided to the WG by the caBIG<sup>®</sup> program, the caBIG<sup>®</sup> tools are used in almost all of the NCI-designated Cancer Centers and caGRID is now the most extensive biomedical network in the US, with more than 145 “nodes.”

## FINDINGS

caBIG<sup>®</sup>'s original mission and goals were important, sound and universally accepted. The program has had a positive influence on the creation and management of standards for data exchange and integration and in supporting computational and integrative tools for basic and translational research. However, a strong consensus emerged from the interviews that, over the past seven years, the program has significantly deviated from that mission and goals. The caBIG<sup>®</sup> program has grown rapidly without adequate prioritization or a cost-effective business model, has taken on a continuously increasing and unsustainable portfolio of development and support activities, and has not gained sufficient traction in supporting critical cancer research community needs. The WG interviews indicate that the program has developed some extremely

expensive software solutions that have not been adopted in a meaningful way by the NCI-designated Cancer Centers, have competed unnecessarily with existing solutions produced by industry leaders that hold a 60% to 70% market share in the NCI-designated Cancer Centers, and ultimately have created an enormous long-term maintenance, administration, and deployment load for the NCI that is financially unsustainable.

In the following sections, we will first address the creation and management of standards for data exchange and support for community-driven tools, the areas in which there was consensus among those interviewed that the caBIG<sup>®</sup> program has been a success. We will then comment on the critical findings in several other key areas of the program, including the impact and track record of the Life Science/Integrative Research Tools, the Clinical Trial Management System, and the Research Infrastructure Tools; the caBIG<sup>®</sup> efforts at outreach and community engagement; and the caBIG<sup>®</sup> program administration and management and budget.

### **Creation and Management of Standards for Data Exchange and Support of Community-based software tools**

The caBIG<sup>®</sup> program has had its greatest impact on cancer research in these areas. In particular, the program has catalyzed progress in the following three areas:

#### **Development of community-driven standards for data exchange and interoperability**

One of caBIG<sup>®</sup>'s greatest successes has been building a consensus among stakeholders for the need to overcome traditional organizational boundaries to enable multi-organizational data, information, and knowledge sharing. Many of the interviewed parties suggested that caBIG<sup>®</sup> has done a great job “putting the right parties around a table” and establishing the relevance of working together to set standards for data exchange and integration. For instance, through its caDSR effort, caBIG<sup>®</sup> has driven the use of community-defined vocabularies. These are now incorporated into the majority of the electronic clinical trial management tools used by virtually all cancer centers, including commercial solutions (e.g. Oracle Clinical or Velos) as well as custom systems developed in house by several larger cancer centers.

On the discovery/research infrastructure side, caBIG<sup>®</sup> has spearheaded the practical implementation of interoperability, at least for cancer research tools. This concept, albeit not new, has been heralded as a critical direction for the development of the next generation of research software tools. Interoperability and common vocabularies are critically important goals that should be supported by the NCI and more broadly by the NIH.

#### **Support for the development, maintenance, enhancement and dissemination of software tools developed by the academic research community**

The caBIG<sup>®</sup> program has provided significant support for academically-developed software tools in areas such as (a) biomedical data storage and retrieval/data mining (e.g., tissues, images, genomic, and proteomic data); (b) integrative genomics; (c) systems biology and computer modeling; and (d) visualization tools. Due to the limited size of the market, these areas have typically lacked commercial solutions and are, as a result, those in strongest need of systematic

and programmatic support by the NCI. Academic software in these areas is often poorly supported, poorly disseminated, and lacks critical support for installation, training, and documentation.

caTissue, which was developed as a collaboration among researchers at several academic centers, GenePattern, a collaboration led by the Broad Institute, and caImage, developed by CBIIT, received almost uniformly positive reviews by the investigators interviewed. caBIG<sup>®</sup> facilitated broad adoption by the community by supporting bug-fixing, dissemination infrastructure, and training materials. Without caBIG<sup>®</sup> support, all but the most computationally sophisticated labs involved in cancer research would have faced a very steep implementation and learning curve that would have prevented their adoption of key tools. The caBIG<sup>®</sup> Knowledge Centers that support these tools are heavily used by the cancer research communities and provide critically needed support for deployment, training, and maintenance of these tools.

### **Establishment of community dialog on interoperability of clinical and research software tools**

Pre-caBIG<sup>®</sup> efforts to create and adopt common standards for data exchange and tool interoperability without funding support mainly created small communities, each driving its own standards and fighting for dominance. The caBIG<sup>®</sup> program has increased awareness throughout the biomedical research community of the need for and potential benefits of service oriented architectures, semantic interoperability platforms, and re-usable research data management tools.

caBIG<sup>®</sup>'s comprehensive program of teleconferences and in-person meetings/workshops has facilitated the creation of an open-source community for the development and adoption/adaptation of extensible biomedical informatics platforms and solutions, with associated communication and collaboration mechanisms and best-practices. caBIG<sup>®</sup> has been able to bring researchers from non-profit and for-profit organizations to the table to discuss how IT and computational approaches can help increase the pace at which data and metadata is generated and managed. While the dialog is far from being completely cohesive or universally accepted, this is a useful activity that should continue under the leadership of the NCI (for cancer) and of other NIH institutes for other disease and basic science related research areas. In particular, the caTissue user community has been widening with clearly evident user feedback and modifications of this tool for local applications.

## **IMPACT AND TRACK RECORD OF caBIG<sup>®</sup> INITIATIVES AND TOOLS**

### **Life Science/Integrative Cancer Research Tools**

The Life Sciences Domain of the caBIG<sup>®</sup> program develops integrative cancer research workspace (ICRW) and informatics tools to support basic and translational research that caBIG<sup>®</sup> has divided into three main areas: (a) Bench-to-Bench research tools that facilitate data exploration, cross domain integration, biomarker selection and quantification; (b) Biospecimen Management tools that facilitate specimen inventory, tracking, annotation and retrieval; (c) Bench-to-Bedside tools that link clinical outcomes with molecular findings.

In area (a), there are 32 Bench-to-Bench research tools (Appendix G) with development costs ranging from \$100K (e.g., Reactome) to \$9.3 M (caARRAY) according to information provided to the WG by caBIG<sup>®</sup> leadership. The caTissue suite, which cost approximately \$5M to develop, is the main caBIG<sup>®</sup> tool in the Biospecimen management area. In area (c), the caB2B (Bench-to-Bedside) tool, which cost approximately \$2.4M to develop, was designed to support query of caGrid for analytical and data services.

The WG found the adoption, track record and impact of the caBIG<sup>®</sup> Life Science/Integrative Cancer Research tools to be quite uneven. Of the 32 Bench-to-Bench tools, the WG received the largest number of positive comments about GenePattern, which was developed by the Broad Institute and made caBIG<sup>®</sup>-compatible with NCI support (\$1.8M). caBIO (\$2M) and caIntegrator (\$1.1M) are used in a few instances, such as TCGA labs.

Several smaller cancer centers have adopted caARRAY, and the caBIG<sup>®</sup> Molecular Analysis Tools Knowledge Center reports about a dozen serious users of caARRAY. However, most of the groups interviewed by the WG who are actively involved in genome research, including those participating in the NCI TCGA projects, are not using caARRAY. The principal reasons given for not adopting caARRAY include: (1) it is too complex and cumbersome; (2) it is difficult to customize; (3) other software tools with comparable capabilities are easier to use; and (4) it will soon be out of date because of the move away from microarray analysis towards RNA sequencing. Several bioinformatics core directors and caBIG<sup>®</sup> deployment leads interviewed reported that they had installed, demonstrated and tried to “sell” caARRAY to potential users in their institution without success. On the whole, therefore, it appears that caARRAY has had limited usage and impact in the cancer centers and among participants in several of NCI’s genomics-based initiatives, which is particularly disappointing in light of its high development cost.

The majority of the 32 Bench-to-Bench research tools developed by caBIG<sup>®</sup> under contracts with commercial or academic investigators have had very limited usage and, as a result, have not generated significant impact in the scientific community. The main factors cited during the interviews were:

- The tools have been re-engineered too many times over the course of the caBIG<sup>®</sup> program.
- The tools tend to be over-designed and overly ambitious, so they cannot be adopted off-the-shelf as “promised” in the early marketing of caBIG<sup>®</sup>. Instead, a significant level of technical knowledge and dedicated local informatics resources are required to make the tools useful in a cancer center’s research environment and to support their customization, adoption and use. This is very frustrating to potential users.
- There is generally inadequate technical support and documentation from caBIG<sup>®</sup>. Questions posted to caBIG<sup>®</sup> tend to be passed down the line, making it difficult for users to solve problems in a timely fashion.
- Many other commercial and open-source software products provide a less severe learning curve for adoption, making them more attractive and cost-effective for most cancer centers.

The caTissue suite, in the Biospecimen Management area, is by far the most widely adopted caBIG<sup>®</sup> Life Science tool. This tool was developed by contracts to the University of Pittsburgh Medical Center and Washington University of St. Louis. caTissue permits users to enter and retrieve data on biospecimens, and it has been evaluated and adopted by many cancer centers and other NCI-supported programs. However, many of the adoptions required significant modifications and local adjustments of the prototype provided by caBIG<sup>®</sup>. Several of those interviewed by the WG reported that caTissue facilitated consolidation of several smaller repositories, allowed for more consistent adherence to IRB protocols, and enabled a more proactively pursued collection of fresh tissue samples for ongoing molecular studies.

The wider adoption and usage of caTissue was facilitated by several factors, including:

- An NCI mandate and supplemental funding to promote the usage of caBIG<sup>®</sup> tools by the NCI-designated cancer centers
- The limited availability of commercial software in this arena
- Two caBIG<sup>®</sup>-sponsored user meetings that were focused on the deployment of caTissue
- The perception by several of the cancer centers that caTissue addressed a non-mission critical need for which they did not have a long-standing legacy system in place

Even so, the interviews revealed several significant barriers to adoption of caTissue by other cancer centers and research programs. Some institutions evaluated caTissue and found it to be not user friendly or lacking in necessary functionality; these institutions went on to develop and implement custom web-based applications to support their biorepositories or had to invest considerable resources to bring caTissue on board as NCI required but also spend quite a bit more for the capabilities they needed. Other reasons cited were that caTissue can only be trusted with rigorously de-identified data, and that it does not support the CHTN standard for non-cancer repositories.

Overall, therefore, the WG interviews reveal that the caBIG<sup>®</sup> efforts in the Life Sciences domain have had a rather limited impact on cancer research across the NCI-supported centers and programs. It is noteworthy that the more widely adopted Life Sciences tools have their roots in projects that were already fairly successfully developed by academic research institutions, whereas most of the caBIG<sup>®</sup>-initiated projects have been less successful and, ironically, were more costly. For example, GenePattern was already well established at the Broad Institute when they received caBIG<sup>®</sup> funding and Reactome, which is a large project with multiple funding sources, was not developed as a tool within caBIG<sup>®</sup>. While not all of the tools developed by academia were successful, it appears that none of the caBIG<sup>®</sup> vendor-developed tools are competitive with comparable commercial software provided by the for-profit sector. Several of the caBIG<sup>®</sup> tools appear to have been developed in response to specific requests from narrowly focused research programs, which pre-destined their limited adoption. Furthermore, because of the limited usage of caBIG<sup>®</sup> research tools, NCI has not yet achieved its original goal for caGrid as the platform of interoperability, where research data can be integrated, standardized and shared across the NCI-funded centers and programs. The difficulties users commonly experienced in incorporating caGRID and the various caBIG<sup>®</sup> tools into existing informatics systems have compromised the caBIG<sup>®</sup> mission as the facilitator of interoperability and data sharing across the cancer research community

## Clinical Trials Management System

The caBIG<sup>®</sup> Clinical Trials Management Systems (CTMS) Workspace has produced a mix of applications developed at various academic and commercial sites, and offers compatibility with commercial offerings, such as Oracle Clinical. There was early, high-level consensus that clinical research support was of the utmost priority for the cancer research community, and clinical data management tools and databases and staff support for them were the two dominant priorities articulated by the 49 cancer centers initially surveyed by the caBIG<sup>®</sup> program in 2003. However, the development of clinical trial management tools by caBIG<sup>®</sup> has lagged significantly behind its tools for bio-specimen and bio-molecular data management.

The caBIG<sup>®</sup> vision was to create a CTMS consisting of an interoperating set of academic and commercial tools that was standards-based to allow the tools to seamlessly interoperate with each other both within and between institutions. This would facilitate the creation of local and group cooperative trials, allow researchers to seamlessly communicate information to the NCI and other regulatory agencies (e.g., FDA), and allow clinical research centers great flexibility in creating and customizing their systems. The potential cost savings of adopting an open source tool, as well as the potential for interoperability, helped win institutional support to participate in the development of the caBIG<sup>®</sup> CTMS. In particular, several of the cancer center directors interviewed indicated that prior to caBIG<sup>®</sup>, there was limited institutional support for making the outlays needed to either develop or buy advanced informatics tools to manage clinical trials.

There were discrepancies in the data provided to the WG by the caBIG<sup>®</sup> program on usage of the caBIG<sup>®</sup> clinical data management tools. One table indicated that only 8 of the 51 cancer centers using caBIG<sup>®</sup> tools had caBIG<sup>®</sup> clinical data management tools in production (Appendix C), but another table showed that 15 NCI-designated Cancer Centers and 8 other extramural institutions were using the C3D/OpenClinica caBIG<sup>®</sup> tools for clinical data collection, management and tracking. Furthermore, clarifications and definitions requested by the WG following the meeting with caBIG<sup>®</sup> leadership on November 2, 2010, also suggest that “accession” data typically reported in presentations about caBIG<sup>®</sup> CTMS and infrastructure tools may significantly overstate actual usage of the caBIG<sup>®</sup> tools.

Based on interviews conducted by the WG, it is clear that the caBIG<sup>®</sup> clinical tools generally lack traction among the Cancer Centers and the broader cancer clinical trials community, and there is minimal community-wide use of patient registries and data capture tools created by the caBIG<sup>®</sup> program. While expressing consistent strong support for the original aims of the CTMS project, only a few of the NCI-designated comprehensive and clinical cancer centers have actually adopted CTMS, and most adopters have selected only a few tools rather than the complete set of offerings. These results also follow the trends noted in 2008 - 2009 in which RedCap, OnCore and Velos were the dominant clinical trials data management systems and only 5 of the 37 institutions reporting use of clinical trials management software at that time reported use of caBIG<sup>®</sup> tools (1). It was particularly interesting that of the 15 cancer centers listed as CTMS developers, adopters and Workspace Working Group members in 2004 (7), only three currently have caBIG<sup>®</sup> CTMS tools in production according to the chart provided by the caBIG<sup>®</sup> program (Appendix C) and three are not listed on the chart at all, suggesting that they are not using any caBIG<sup>®</sup> tools.



Several of the potential adopters interviewed had performed extensive testing of the CTMS system and ultimately either chose a competing commercial product or elected to continue to develop home-grown alternatives. When pressed to explain their decisions, the reasons most commonly cited were that the CTMS tools were incomplete, too generic, and/or overly complex, and that the user interfaces would require extensive and expensive customization in order to make them user-friendly. In addition, some of the caBIG<sup>®</sup> clinical research management tools, specifically caAERS, C3PR, and the Patient Study Calendar, were said to be highly error prone (“buggy”) and require complex technical and workflow modifications to adequately meet the needs of users. Those interviewed also felt that some of the modules were “over-sold” and promised more than they actually delivered. Other commonly-cited reasons not to adopt CTMS include:

- The existence of critical legacy systems that would require deep and costly integration with any comprehensive clinical trial management software solution. Unlike commercial vendors, caBIG<sup>®</sup> does not offer software developer resources to perform this integration. For smaller cancer centers, this is a major issue.
- The larger cancer centers have, in many cases, already made multi-million dollar investments in electronic clinical trial management tools and in their integration, and have little motivation to switch to new solutions.
- The CTMS tools are insufficiently modular, creating unwanted interdependencies. For example, the Patient Calendaring module must be present even if it is not needed. This increases the cost to maintain the installed software and reduces its flexibility.
- The inconsistent user interfaces and awkward linkages among different CTMS modules, with missing or incompletely implemented features.

In addition, the NIH CTSA program has introduced uncertainty about which data standards will ultimately be adopted. This was cited by several interviewees as a reason for proceeding with caution in adopting CTMS.

Therefore, despite many years of development effort at a cost of at least \$50M, the adoption of clinical trials management tools by the clinical and translational research community has increasingly gravitated away from caBIG<sup>®</sup> technologies. It is noteworthy, however, that several commercially available software products have partially adopted caBIG<sup>®</sup> common vocabularies and data standards; for example, Oracle Clinical interoperates with various CTMS tools and OnCore has recently become caBIG<sup>®</sup> compliant at the bronze level.

Representative comments from interviewees include:

- We already do this quite well with OnCore, and I was mainly interested in using the tool for its reporting capabilities to AdEERS or MedWatch or both, as an add-in with OnCore obtaining needed help with Forte Research (PercipEnz). However, the next release of OnCore this spring will have the ability to automatically report out to both AdEERS and MedWatch directly. They have done this by working directly with these organizations. The use of caAERS as a middleware piece then becomes unnecessary.
- CTRP is supposed to register protocols and eliminate duplication of effort for registration in clinicaltrials.gov (required by law). However, because of a timing issue, CTRP system can't be used for this and duplicate registrations are still necessary.

- What caBIG<sup>®</sup> created was a series of modules that required different software stacks and weren't completely interoperable. No real central database; we would have to provide our own or use C3D, which had its own issues.

On a brighter note, work on a standard set of Case Report Forms (CRFs) in the CTMS workspace has had some success. This effort has been supported by CDISC and caBIG<sup>®</sup> and represents a collaboration between industry, the NCI cooperative clinical trials groups, and others performing cancer clinical trials. There were also a few reports of caBIG<sup>®</sup> CTMS tools facilitating specific clinical trials, including the I-SPY trial with an adaptive trial design and one between Duke and investigators in China.

Overall, therefore, it appears that the caBIG<sup>®</sup> program has had a positive impact on creating common data models and other standards for clinical data management, but has not met expectations for providing software systems for use by end users. Finally, we were unable to identify any adopter who was taking advantage of CTMS's ability to connect clinical trials databases to the grid, or to seamlessly transmit adverse events and other reportable information to regulatory agencies.

The WG is therefore concerned that, on its current path, the caBIG<sup>®</sup> clinical program will have a rather limited impact on cancer clinical trials.

### **Infrastructure Tools**

caBIG<sup>®</sup>'s infrastructure activities can be grouped into three broad categories: (1) the development of standard vocabularies and data models; (2) caBIG<sup>®</sup> imaging tools; and (3) the caGRID system, which provides secure interprocess communications for sharing of data and computation among caBIG<sup>®</sup> systems. The success of these infrastructure activities has been mixed.

The development of data standards and an ontological framework that allows data to be harmonized across cancer centers is widely perceived to be one of the most important contributions of the caBIG<sup>®</sup> program. The creation of the caDSR (Cancer Data Standards Registry and Repository) and the EVS (Enterprise Vocabulary System) provided new resources that could be used by cancer centers and software developers, academic and commercial, to begin creating a framework for integrating data from heterogeneous sources into usable datasets. This infrastructure is being used beyond caBIG<sup>®</sup> tools and infrastructure and has increased awareness throughout the cancer community that data needs to be collected, stored and used in a common way such that multiple researchers and research teams can work together on common projects. Having said that, as with other aspects of the caBIG<sup>®</sup> program, there were less positive comments about the way the standards development had been carried out. Those interviewed commented that the process is driven by technologists who do not sufficiently engage with researchers; that tools for inputting data such as for caDSR have steep learning curves and are laborious to use; and that the caBIG<sup>®</sup> process often takes much too long to release a standard and ignores the fact that well developed standards may already be in use by many groups.

caBIG<sup>®</sup> imaging tools and workspace have had some impact, though with similar qualifications. Several institutes are exploring or have adopted the NBIA (National BioMedical Imaging

Archive) for images. The technologies being developed in the caBIG<sup>®</sup> imaging workspace (especially the Annotation and Image Markup standards, "AIM") for image metadata are also being adopted by the vendor community in their commercial workstations. However, users found the caBIG<sup>®</sup> imaging tools difficult to install, hard to populate since they are not well integrated with other systems, and inadequately security tested.

In comparison, the adoption of caGRID and related infrastructure software has been very limited. A great deal of time, effort and resources (at least \$8.9M) have been expended creating caGRID. According to the chart provided to the WG by the caBIG<sup>®</sup> program (Appendix C), only 7 of the 51 cancer centers using caBIG<sup>®</sup> tools have caBIG<sup>®</sup> infrastructure tools in production, 20 are "in process", 2 are planning implementation, and 21 have no activity or plans to implement these tools. It is interesting that the University of Chicago, which was involved in developing caGRID, reports not using it. The WG members interviewed a large number of IT directors and other individuals directly involved in testing or deploying caGRID at their institutions. Their comments indicate that caGRID has been a very difficult technology to implement, maintain and use. The main stated barriers to adoption are:

1. caGRID is quite complex and requires resident experts in Java programming to support the data integration both across the center and for institution to institution communication.
2. There is no graphical user interface to simplify basic administrative or configurational tasks.
3. Constant changes in the grid architecture and individual tools ("software churn") increased barriers to adoption and made commercial offerings more attractive, even if they did not offer the same promise of data sharing and common semantics.

Beyond these considerations, several individuals interviewed noted that caBIG<sup>®</sup> has not adequately harmonized caGRID and its associated software tools into a single architecture, and that caBIG<sup>®</sup> software products that are nominally caGRID-enabled do not in fact interoperate well with other grid-enabled tools. In addition, the currently planned full rewrite of the caBIG<sup>®</sup> infrastructure has created considerable uncertainty in the cancer centers and funding to cancer centers to assist in adoption has decreased.

The TCGA analysis centers must make their tools "caBIG<sup>®</sup> compliant"; however, interviewees said it is relatively easy to make them "bronze level" compliant, which means they have documented web service APIs. Several senior investigators involved with the TCGA project were not aware of any use of their data on the caBIG<sup>®</sup> grid, and their discussions with other investigators during a session on interoperability at a recent meeting of the TCGA analysis groups in Long Beach, California, confirmed that few if any of the investigators participating in TCGA data analysis were taking advantage of the caBIG<sup>®</sup> grid for data sharing.

The following are representative comments about caGRID:

- caGrid is of very limited use. Part of the problem is that not much data is currently being shared there, and part is the complexity and cumbersomeness of the system design.
- caGRID - implemented but not used. Cancer center doesn't really want to share data anyways
- Not rigorously security tested

Those cancer centers that did have projects involving data sharing found it easier to use other systems rather than caGRID.

## **caBIG<sup>®</sup> Community Engagement**

The caBIG<sup>®</sup> program has had an aggressive community engagement effort which includes:

- Annual caBIG<sup>®</sup> meetings
- An extensive caBIG<sup>®</sup> Website
- Monthly conference calls for each of the Workspaces and Special Interest Groups
- Five Knowledge Centers which provide technical assistance with the Tissue/Biospecimen Banking and Technology Tools, caGRID, the CTMS tools, Data Sharing and Intellectual Capital issues, and the Molecular Analysis Tools
- A caBIG<sup>®</sup> Deployment Lead (DL) program which provided administrative supplements to NCI-Designated Cancer Centers for support of dedicated staff to facilitate awareness and adoption of caBIG<sup>®</sup> tools
- caBIG<sup>®</sup> demonstrations and exhibit booths at many cancer research meetings
- caBIG<sup>®</sup> newsletters
- Periodic caBIG<sup>®</sup> Summits for strategic planning

Comments received during the interviews indicate that these outreach efforts have generated mixed reactions from the cancer research community. Most of the cancer centers interviewed were glad to receive the additional staffing resources provided through the caBIG<sup>®</sup> DL program. Several of those interviewed commented that the DLs kept them aware of developments in the caBIG<sup>®</sup> program even though the center ultimately decided not to use caBIG<sup>®</sup> tools. However, the quality of the conference calls for the DLs was reported to be very uneven.

The Knowledge Centers also generally received praise, and the caBIG<sup>®</sup> workgroups have evolved and matured over the years and are now generally viewed as very productive. Although the caBIG<sup>®</sup> website is very extensive, there were also several comments that there is obsolete and “buggy stuff” on the caBIG<sup>®</sup> website (such as a virtual appliance that contains the complete caTissue application that does not work), and that some of the pages could be organized better. There were also comments that caBIG<sup>®</sup> was slow to recognize the potential of social media tools and actively discouraged the use of tools such as WIKIs.

Although a few of those interviewed found the annual caBIG<sup>®</sup> meeting useful to pull together issues, allow opportunities to network, and focus cancer center directors on informatics, most of those interviewed found the annual meetings frustrating because the meetings have been more focused on marketing and growing caBIG<sup>®</sup> than providing concrete plans for solving real-life problems faced in the cancer centers and making the caBIG<sup>®</sup> tools already available work better for them. For example, the 2010 summer caBIG<sup>®</sup> Jamboree was described as a “contractor feeding frenzy” with “a bunch of people in suits hoping to get a piece of the pie.”

The caBIG<sup>®</sup> program has also had several meetings specifically for strategic planning and gathering of input from the community and has engaged the MITRE Corporation to assist with strategic planning. However, the strong perception among those interviewed is that the strategic

plans were aimed at growing the program, and that internal decision-making was not really influenced by comments provided at those meetings. A typical comment was that “feedback to caBIG<sup>®</sup> occurs, but nothing much seems to change”.

### **caBIG<sup>®</sup> Program Administration, Contracts Management and Budget**

The caBIG<sup>®</sup> program is administered by the NCI’s Center for Biomedical Informatics and Information Technology (CBIIT). Information from the caBIG<sup>®</sup> program identified about a dozen CBIIT senior staff who direct various aspects of the program and appear to form the caBIG<sup>®</sup> Program Oversight Board. The latter is in charge of making final decisions about the direction of the program (7). Surprisingly, the Program Oversight Board does not include external members and is not supported by any independent domain expert advisory group, such as a Scientific Advisory Board.

Program operations are managed and coordinated through a variety of contract and subcontract procurement mechanisms. Specifically, program management activities are supported by a large program management office contract to Booz-Allen Hamilton (BAH) for the support of CBIIT Federal staff. Similarly, user community and technical operations, including software development, are supported by a large contract to SAIC Frederick (SAIC-F), the contractor that manages NCI’s FFRDC in Frederick, MD. These two contractors also advise the Program Oversight Board. Finally, individual caBIG<sup>®</sup> projects, tool development and maintenance, Workspaces, Knowledge Centers, In Silico Centers, and other activities are supported through specific subcontracts issued by either BAH or SAIC-F. This results in a multi-layered and very complex organizational structure and lines of authority and responsibility (Appendix H).

To start a new subprogram or contract, the designated main contractor (BAH or SAIC-F) first issues a Request for Proposals (RFP), then evaluates the resulting proposals using an internal review process that is not based on traditional NIH peer review standards, and finally selects awardees to perform the required work based on a Statement of Work (SOW). Payments from BAH or SAIC-F to the subcontract awardees are based on achieving specific predefined deliverables as per the SOW.

During the presentations to the WG on November 2, 2010, it was explained that this subcontracting process for managing the caBIG<sup>®</sup> program was chosen to be “nimble” and to “react in real time to changing needs in the community”. Interestingly, the subcontracting process allows the caBIG<sup>®</sup> program to start new projects and initiatives rapidly because it does not require the standard NCI concept review and approval process. However, avoiding concept review prevented the normal dialog that helps guide NCI to invest in programs that will be the most valuable to the cancer research community.

The information provided to the WG by the caBIG<sup>®</sup> program about its budget and expenditures, especially the ARRA money, was inconsistent and extremely difficult to decipher. As a result, it is not clear exactly how much was spent on the development of each caBIG<sup>®</sup> tool or on program overhead costs.

There was strong consensus from the interviews on the following points:

**1. The caBIG<sup>®</sup> program management structure is overly complex and expensive.** According to budget information provided by the caBIG<sup>®</sup> program, overhead for program and contract management accounts for 25-30% of the total caBIG<sup>®</sup> program expenditures. During the presentations on November 2, 2010, caBIG<sup>®</sup> program leaders accepted this very high overhead rate as inevitable. Overall Program Management overhead for Fiscal Years 2004 – 2010, not including the ARRA projects, for the three main caBIG<sup>®</sup> contractors (BAH, SAIC-F, and Sapient) has cost at least \$60M and probably significantly more. If these funds had been distributed to NCI-Designated Cancer Centers, each center could have received an additional \$200K/year for 5 years to support software license costs. Interestingly, only a fraction of this overhead cost has gone to support tools that have been adopted by more than a handful of labs.

In addition, the third party management of caBIG<sup>®</sup> projects has introduced “too many management layers between the community doing the work and the caBIG<sup>®</sup> leadership,” thus creating a “disconnect” between the original goals of the program and the daily directives that have been used to manage the projects. Interviewed Cancer Center representatives told the WG members that communication and input between the centers and the NCI has been dramatically hampered by the contractor interface and that they sometimes received conflicting directions from the various management layers.

**2. The contractors did not really understand the cancer research space in which they were operating.** There is a strong perception that the contractors steered funding to address the problems they were excited about, rather than addressing the real needs of the cancer research community. Since they were experts in technology but not in cancer research, the program ended up supporting massive new technology developments, such as caGRID, which has had very little impact on cancer patients and cancer researchers. Many researchers felt that “something went horribly wrong” and that the contractor management team was making key decisions about tool development which they were not scientifically and technically qualified to make. Not surprisingly, there appear to be no obvious examples in sight where availability of this technology has made or could make a clear contribution to addressing a cancer relevant problem. Most individuals interviewed were “generally impressed with the caBIG<sup>®</sup> management team’s ideas” and vision but could not say the same about the contractors.

The WG is very concerned that this trend is continuing in the expensive and expansive caBIG<sup>®</sup> projects currently underway. There is insufficient engagement of both informatics scientists and basic/clinical/translational researchers in the design, implementation, and evaluation of caBIG<sup>®</sup> technologies and platforms. For example, the current caGrid 2.0 Roadmap efforts, which are charting the future direction for core technology development in the caBIG<sup>®</sup> program, are led by technical architects, project managers, and software developers from CBIIT and affiliated public and private sector contractors. The roadmap, goals and timeline are opaque to the community. There is almost no participation in this process, other than during public comment periods, by thought leaders in the informatics, scientific, and clinical communities.

**3. The internal processes for soliciting and evaluating proposals for subcontracts are not transparent.** While the RFP management process was (on paper) transparent, the evaluation criteria and the engagement of the community have been completely ad hoc, with funding decisions based not on established peer review criteria but rather on the internal

decisions of the caBIG<sup>®</sup> management team. The community has been generally frustrated by the short turn-around time between issuance of RFPs by the primary caBIG<sup>®</sup> contractors and the proposal due date; the long intervals of uncertainty between review, selection and award of subcontracts for particular tasks; the frequent lapses in subcontracts; and the short duration of many of the subcontracts for research to be done in academic centers.

Both academic and industrial groups alike questioned the rationale and criteria for selecting caBIG<sup>®</sup> awardees. For instance, Velos, Forte, cooperative group representatives, and others raised the issue of the Medidata contract award as part of the CTMS, which created tremendous friction and resulted in an open investigation and eventually in the cancellation of the Medidata contract. Specifically, rather than selecting Electronic Data Capture (EDC) module(s) with an established user base in cancer research, the decision was made to select Medidata, based on the company's willingness to provide Open Source code, which allowed integration with caBIG<sup>®</sup>'s own software for clinical trial data management. Despite its quality, this product has very few adopters among cancer centers. Thus, while Open Source is a valuable principle, this choice was doubly inefficient by (a) attempting to force adoption of an EDC module that was not an industry standard in cancer and (b) making a decision based on a limited rationale rather than the overall benefit to the cancer clinical research community. The effect of such an unpopular choice has been that the availability of a robust solution for electronic data capture has been significantly delayed.

**4. Participation of the same contractors in both program management and software development has the potential for conflict of interest and unfair competitive advantage.**

This situation is particularly an issue regarding caBIG<sup>®</sup> "compliance levels" for software. Industry representatives pointed out that no industrial software can achieve better than "Bronze level" caBIG<sup>®</sup> compliance, because there is not a defined process for certifying software at higher compliance levels. However, internally developed caBIG<sup>®</sup> software, such as caTissue, is automatically labeled as silver compliant because it was designed by caBIG<sup>®</sup> according to its standards, even though it did not receive formal certification. This situation makes it harder for non-NCI organizations to compete fairly and is not conducive to the establishment of high-quality, broadly adopted standards.

**5. There is a perception that caBIG<sup>®</sup> favors an "in group" of participants.** Initially the caBIG<sup>®</sup> program attempted to be all-inclusive. Rather than creating a small community of technically proficient investigators, it tried to create the largest possible consensus across the entire community of cancer center researchers. This was important and valuable, as the process to define common standards is as much an exercise in social networking as it is in technology development. After this initial phase, however, a smaller community of funded labs has emerged and the perception in the broader research community is that it is difficult for other investigators to enter the program. This would not be a problem if the decisions were made based on peer review of proposals. Unfortunately, the decision process has been almost completely driven by the caBIG<sup>®</sup> management committee, resulting in choices that appeared arbitrary and not necessarily optimal to many investigators that participated in the subcontracting process. Additionally, many investigators feel that their voice is being increasingly ignored, thus further opening a chasm between the base of potential users of the caBIG<sup>®</sup> tools and the program management team.

## CONCLUSIONS

Intense scrutiny of the caBIG<sup>®</sup> program, through a large number of community interviews and review of relevant documents, has revealed a relatively clear and picture of the current state of the caBIG<sup>®</sup> program, and its main strengths and deficiencies. The WG assessment process has also provided important information that will be useful not only in reshaping and refocusing this program but also in thinking about future large-scale NCI initiatives and programs.

There was strong consensus about the value and timeliness of the original vision for the caBIG<sup>®</sup> program, which was to define the development of common standards for data sharing and interoperability and to support community-based software tools for the collection, management and analysis of cancer related data as a mission critical requirement for the NCI. These are the areas where caBIG<sup>®</sup> was most successful and made its most relevant contributions. In particular, caBIG<sup>®</sup> succeeded in bringing a vast and diverse community of cancer researchers and practitioners together, through a complex network of workgroups and initiatives, to define standards for data collection, exchange, and interoperability. Equally important has been caBIG<sup>®</sup> support for academic, community-driven software for the analysis of cancer data. Nevertheless, it is noteworthy and a lesson for the future that the more widely adopted caBIG<sup>®</sup> tools have their roots in projects that were already successfully developed by academic research institutions (e.g., GenePattern) or where the community was coalesced by caBIG<sup>®</sup> around an unmet need (e.g. caTissue), whereas most of the caBIG<sup>®</sup>-initiated projects have been less successful and, ironically, far more costly.

However, caBIG<sup>®</sup>'s successes have come at a tremendous financial burden and are offset by several critical deficiencies in the program. caBIG<sup>®</sup> has expanded dramatically in the last few years, transforming into a hugely complex enterprise for the production of NCI-branded software tools, consisting of disparate and loosely connected activities supported by NCI staff, contractors, users, workspaces, and other entities. The WG believes caBIG<sup>®</sup>'s limited overall traction in the cancer research community is due to several fundamental problems in the approaches used to implement the caBIG<sup>®</sup> program, including:

**1. A “cart-before-the-horse” overly broad grand vision for the program.** As discussed, caBIG's original vision was both appealing and shared by many. Unfortunately, that vision was developed and implemented by a technology-driven approach. Rather than identifying pipelines and applications that could revolutionize cancer research and supporting development of the technologies necessary for their implementation, the caBIG<sup>®</sup> program developed of an entire array of ambitious and costly technologies (e.g., caGRID, CTMS, caArray, etc.) with the expectation that these would eventually be integrated into cancer-critical pipelines and applications. Most of the cancer centers interviewed have the perception that tools were developed in a vacuum without sufficient attention to how they would be eventually used and adopted by the community. To a large extent, the In Silico Research Center program was started precisely to fix this problem, and to create a visible, flagship effort that could demonstrate cancer-relevant impact with existing caBIG<sup>®</sup> tools or to set the stage for the next wave of cancer-relevant tools. Instead, this type of activity (as well as similar initiatives in clinical and translational research) should have provided the initial catalyst and motivation to develop the specific technology components necessary to deliver the next generation of cancer-relevant applications.



In addition, caBIG<sup>®</sup>'s technology-centric approach has not adequately addressed key cultural and legislative barriers to the exchange of clinical data, such as firewalls associated with institutional IRBs. The program did not sufficiently define at the outset what "data sharing" meant or exactly what data was needed. Most investigators that have adopted caBIG<sup>®</sup> tools only use them to share data internally, within a lab or a small research group, and have little evident desire to share data across cancer centers/research institutions. Therefore, caBIG<sup>®</sup> is unlikely to be successful anytime soon in achieving its originally intended goal of linking the cancer research community in a single Grid for data sharing. Accomplishing this valuable goal will require the NCI to exercise substantially more "political will", to implement additional more concrete incentives for data sharing, and to adopt a substantially different approach to tool development and dissemination.

**2. A "build it and they will come" mentality.** The idea that availability of functional, open-source, free software for electronic clinical trial data acquisition and management would over time promote its integration in the clinical research enterprise does not sufficiently address the complexities and costs associated with software integration, support, and training by research institutions. Not surprisingly, Cancer Center IT administrators and industry leaders alike have significantly criticized what appeared to be a well-intentioned and possibly groundbreaking initiative by caBIG<sup>®</sup> to support development of free, open source software for clinical data management. This is because the licensing cost for this kind of software pales in comparison to other costs, such as those associated with training the data managers and nurses on a new system, maintaining and updating software over time, and creating/deploying interfaces for its integration with legacy systems for labwork warehouses, Electronic Informed Consent, Electronic Medical Records, and a variety of additional systems routinely integrated in the clinical settings where cancer patients are treated.

Therefore, only a few of the NCI-designated Cancer Centers have adopted the entire caBIG<sup>®</sup> CTMS. Testing of specific components in most cancer centers is relegated to a few pilot projects, mostly funded by caBIG<sup>®</sup>, and cancer center management has expressed little interest in integrating them with their enterprise solutions in the long run. Today, 60% to 70% of the cancer centers use either the Oncore or the Velos electronic clinical trial management systems. When this is contrasted with the ~\$100M in ARRA funds dedicated to the continued development of an NCI-branded electronic clinical trial management software, it is clear that these priorities need to be re-evaluated and re-aligned before the money is spent.

**3. An unfocused attempt to address all problems in clinical and basic research.** The initial caBIG<sup>®</sup> survey of the cancer centers in 2003-2004 clearly showed that the dominant needs of the cancer centers were in clinical data management. However, rather than concentrating on solving a few critical problems well, the caBIG<sup>®</sup> approach has been to tackle virtually every aspect of cancer research, resulting in an extraordinarily ambitious and diffuse program. The opinion of those interviewed is that caBIG<sup>®</sup> has spread itself too thin, thus over-promising and under-delivering, and that it lacks focus, thus frequently delivering solutions to problems that are not the most important ones faced by the cancer research community.

Only a small number of the tools developed by caBIG<sup>®</sup> are deemed to be useful to those interviewed, and even fewer have been widely adopted by the community. This is especially true in the clinical data management arena, where the largest caBIG<sup>®</sup> investment has taken place.

Remarkably, most of the program traction has been in areas where caBIG<sup>®</sup> has made smaller financial investments.

**4. A “one size fits all” approach to funding and management of scientific and software development projects.** Many of those interviewed dislike the opaque, non-peer-reviewed, and inflexible subcontracting process used by caBIG<sup>®</sup> to manage all of its projects. While the acquisition approach may be appropriate to managing caBIG<sup>®</sup>'s software development projects with defined deliverables, it is perceived to be less than ideal for caBIG<sup>®</sup>-supported research projects conducted in the academic community. For instance, timelines to submit proposals to caBIG<sup>®</sup> announcements have often been extremely short, giving the impression that decisions on funding were already made before the proposals were received. After proposal receipt, the timeline for funding was ambiguous and unfunded proposals received inadequate feedback about why they had not been selected. Subcontracts issued to academic institutions are often of short duration, and have frequent lapses, rigid staffing requirements and onerous monthly reporting burdens that are not conducive to academic research. These contracting requirements are not the most effective for development of complex algorithmic tools that are heavily science-driven or for research activities, such as the In Silico centers. Finally, the internal decision-making process is opaque to the community and appears to be based solely on the priorities and interests of the caBIG<sup>®</sup> management team.

The layers of contract management organizations also hampered and discouraged direct communication between caBIG<sup>®</sup>-funded groups and the NCI. This has resulted in an increasing chasm between the original priorities of caBIG<sup>®</sup> and their implementation as envisioned by the contract management organizations. Not only is this expensive for NCI in terms of overhead costs, but it also creates the requirement for academic institutions to hire additional project managers to interface with the contract management organizations and to handle the significant reporting requirements.

**5. A business model that is unsustainable and not cost-effective for the NCI or potential users.** Software development is not NCI's core mission. The caBIG<sup>®</sup> program appears to have adopted a “free-for-all” policy for all software associated with the management of both clinical and basic science cancer research data. While these are laudable goals, it is unclear how the NCI can sustain such a strategy in the long term, as this includes extremely costly maintenance, upgrade, and dissemination requirements. The dedicated personnel and labor intensive support that caBIG<sup>®</sup> has provided to a few cancer centers to promote adoption of the CTMS tools is clearly not scalable across all cancer centers and clinical trials institutions.

In addition, the WG assessment process revealed another potential consequence of this caBIG<sup>®</sup> funding practice. Although the first versions of caBIG<sup>®</sup> were released several years ago, it was notable that many of the concerns raised during interviews with WG members had apparently been largely unspoken. The collective silence about problems with caBIG<sup>®</sup> software and approaches may be at least partially because cancer centers were receiving additional funding through the caBIG<sup>®</sup> DL program to make their tools “caBIG<sup>®</sup> compliant”. In a number of instances, we were told that cancer center staff supported by caBIG<sup>®</sup> spent only part of their time on “caBIG<sup>®</sup> integration” and the rest supporting the cancer center's general informatics activities. This created disincentives for reporting non-performance of caBIG<sup>®</sup> tools.

One of caBIG<sup>®</sup>'s approaches has been to license Service Providers that can be hired to implement and support caBIG<sup>®</sup> tools and infrastructure. However, this approach is also problematic and did not adequately address the many business considerations that cancer centers have when deciding on the informatics tools that will be deployed. Centers typically require extensive, multi-million-dollar efforts to effectively integrate new tools within a pre-existing patchwork of legacy systems and databases. If a set of open-source caBIG<sup>®</sup> tools still requires commercial vendors for installing and maintaining the infrastructure, there is little advantage over commercial tools that are more mature architecturally and far easier for the researchers to use.

Finally, it is not clear that caBIG<sup>®</sup> adequately weighed cost-effectiveness vs brand identity in its decision-making about tool development. Several costly tools, such as caARRAY, which cost more than \$9M to develop according to data provided by the caBIG<sup>®</sup> program, were not widely adopted by the community because of the availability of well established alternatives that were either free or commercially available (e.g., ArrayExpress, GEO, GeneSpring, etc.). A small fraction of the total caARRAY investment would have been sufficient to help implement innovative or critically missing features (e.g., gridification, query interface, etc.) in the more widely used array analysis tools rather than creating a completely new solution from scratch. Although "leveraging of existing resources" is a stated caBIG<sup>®</sup> principle, the perception is that adaptation of existing vendor or in-house tools has not been a major area of interest within caBIG<sup>®</sup>. Software vendors attended many caBIG<sup>®</sup> meetings, but the moving scopes and lack of concrete requirements for many caBIG<sup>®</sup> projects have been disincentives for them.

**6. Development and management of clinical informatics and basic science discovery tools under one umbrella organization.** While it may be convenient to think of clinical informatics and bioinformatics/computational-biology as inter-related disciplines, requiring joint administration, this is not the standard in current practice outside of caBIG<sup>®</sup>. The two communities have different goals, publication venues, metrics for success and, with few exceptions, researchers do not span across the two disciplines. Experts in clinical informatics are unlikely to be the best judges of the value and innovative nature of a new bioinformatics approach and vice versa. It is likely that by consolidating support for these areas under one organization, the NCI has created an artificial division between the communities that are the real producers and consumers of informatics and algorithmic innovation and the program that is meant to address these needs.

The NCI already invests significant efforts in the development of new computational and systems biology methodologies via the Integrative Cancer Biology Program (ICBP) and other programs that aim to develop new methodologies for the integration of computational and experimental approaches. These programs, however, lack funding mechanisms to support key activities, such as software tool hardening, community dissemination, and personnel training, which are necessary for the novel methodologies developed by investigators funded under this program to be broadly adopted by the research community. More importantly, there is very little cross-funding and support from caBIG<sup>®</sup> that goes to support these crucial programs.

**7. Lack of independent scientific oversight of goals, priorities, projects and evaluation of progress.** The caBIG<sup>®</sup> program has held numerous user meetings and engaged MITRE Corporation to assist with strategic planning and process development for expanding the program and fostering adoption of caBIG<sup>®</sup> tools. However, there has been a complete lack of

independent program review and oversight from non-caBIG<sup>®</sup>-associated experts from the very beginning of the project in 2000 – 2002 to the present. For example, there has not been independent scientific review of caBIG<sup>®</sup>'s initiatives or projects at the concept stage, and the program does not have an external Scientific Advisory Board to provide periodic assessment of its vision, objectives, strategies, approaches, priorities, and metrics for success or to advise the caBIG<sup>®</sup> management of the cancer relevance of the program as a whole. Most of the problems in the caBIG<sup>®</sup> approaches listed above might have been avoided if there had been adequate independent external input as the program developed and expanded.

**In summary**, the WG concludes that the massive investments made by the NCI in the caBIG<sup>®</sup> program have not translated into commensurate traction in the cancer research community and, ultimately, into impact on cancer research. As a result, substantial rethinking and restructuring is needed to bring the program back to its original mission and goals, which remain of critical relevance to the NCI mission. Many of those interviewed felt that caBIG has become too focused on its own development. By transforming from a highly competent and strategically motivated enabler of information technologies for cancer research to a one-stop-shop developer of software solutions, caBIG<sup>®</sup> has not delivered on its originally intended goals and has created a financially unsustainable long-term software maintenance, customization, and dissemination situation.

Despite the problems in the caBIG<sup>®</sup> program identified by this review, the WG feels strongly that cancer research has never been as dependent as it is now on electronic and computational capabilities to acquire, store, manage, and analyze large volumes of clinical and genomic data. As a result, continued support of clinical informatics tools and algorithmic advances remains a mission-critical requirement for the NCI. The NCI must establish appropriate measures to encourage and oversee the development of the next generation of tools and algorithms necessary to improve cancer treatment, prevention, and diagnosis.

## RECOMMENDATIONS

Based on our findings, and with the goal of sustaining the positive aspects of the caBIG<sup>®</sup> program, the WG members unanimously agree on several recommendations. The Immediate Tactical Recommendations address critical time-sensitive issues that have emerged from this study. The Longer Term Strategic Recommendations are directed at refocusing, restructuring, and resizing the caBIG<sup>®</sup> program in the context of NCI bioinformatics initiatives as a whole.

### *Immediate Tactical Recommendations*

**1. Institute an immediate moratorium on all ongoing internal and commercial contractor-based software development projects while initiating a mitigation plan to lessen the impact of this moratorium on the cancer research community.** The moratorium should encompass caBIG<sup>®</sup>'s current projects on the enhancement and development of tools in the CTMS suite, the caGRID, and the activities in the caBIG<sup>®</sup> 2.0, cloud computing, electronic health records, research support and population science initiatives. To avoid negative impact on current caBIG<sup>®</sup> users, a mitigation plan should be put in place to provide support for their established mission-critical activities that depend on the availability of specific caBIG<sup>®</sup> tools that may be affected by this moratorium. Based on the results of the interviews, the WG recommends that maintenance

and support for caARRAY, caTissue, the imaging tools, and ongoing multi-center clinical research projects, such as the I-SPY trial, not be subject to the moratorium at this time.

**2. Institute a one-year moratorium on initiation of new projects, contracts and subcontracts by caBIG<sup>®</sup>.** This moratorium should affect all efforts under consideration but which have not yet been published as formal requests for proposals by the caBIG<sup>®</sup> contractors. The moratorium should also apply to existing, already published announcements for which proposals have not yet been received. An independent oversight committee (see Recommendation 4 below) should be charged with recommending, within one year, appropriate future funding mechanisms for projects and initiatives that the committee finds to be meritorious. All future informatics initiatives should go through the standard NCI concept review and approval process (see Recommendation 9 below).

**3. Provide a one-year extension on caBIG<sup>®</sup>-supported academic efforts for development, dissemination, and maintenance of new and existing community-developed software tools.** This interim support will be critical to avoid a deleterious effect on the academic community currently involved in creating solutions to the informatics needs in cancer research. For contracts already awarded to academic research units (but not commercial vendors), continuation should be for the shorter of one year or the original contract expiration date. New proposals that have already been submitted by academic research units may be funded for up to one year based on the established review process. Within one year, the independent oversight committee (see Recommendation 4) should review the existing contracts to academic research units and identify appropriate funding mechanisms to provide continued support for projects deemed to be scientifically meritorious.

**4. Establish an independent oversight committee, representing academic, industrial, and government (NCI, NIH) perspectives to review ongoing and planned initiatives for scientific merit and to recommend effective transition options for current users of caBIG<sup>®</sup> tools:** The committee should determine whether individual tools and technologies developed by caBIG<sup>®</sup> should be discontinued, transferred to the academic or industrial community with appropriate support, or continued internally under a focused clinical and research impact-driven management plan. For instance, existing adopters may be funded to acquire licenses to equivalent commercial software and to achieve an equivalent level of integration within their environment.

**5. Conduct a thorough audit of all aspects of the caBIG<sup>®</sup> budget and expenditures.** Given the complexity of the caBIG<sup>®</sup> budget due to the contracting/subcontracting structure of the program and the inconsistencies noted in the budget data provided to the WG, a thorough audit of its financial transactions is recommended. This audit should be used to identify unspent funds that may be recovered for reprogramming for use in implementing other recommendations, such as the mitigation plan suggested in Recommendation 4 and future funding mechanisms to be recommended by the independent oversight committee, and for other NCI priorities.

### *Longer Term Strategic Recommendations*

**6. Create an independent Scientific Advisory Group (SAG) for NCI biomedical informatics efforts and initiatives that includes scientific, technology and informatics expertise to advise NCI on appropriate informatics priorities, initiatives, business model(s), and resource**

**allocations.** This committee should be tasked with defining the informatics and algorithmic needs of the clinical, translational, and basic science research community, while recognizing that these are not necessarily unique to cancer research. The SAG should facilitate the abatement of barriers with similar efforts in other NIH institutes (e.g. NHGRI), in the community and abroad (e.g., EMBL/EBI and ICGC) to address the current perception of caBIG<sup>®</sup> as an insular and highly balkanized effort. It may be appropriate that this SAG becomes a standing subcommittee of the BSA with appropriate external help from additional highly qualified domain experts.

**7. Refocus caBIG<sup>®</sup> on its original mission and discontinue all strategic efforts to develop and maintain its own brand of software tools, either directly or indirectly through commercial contractor efforts.** In the future, caBIG<sup>®</sup> should focus exclusively on (a) helping define standards for interoperability and data exchange, (b) working with the academic and for-profit communities to facilitate the integration and adoption of these standards into clinical and basic science research software, and (c) supporting valuable academic software tools that have a proven track record of scientific innovation in cancer research.

**8. Separate the clinical informatics and bioinformatics components of the caBIG<sup>®</sup> program.** Specifically, efforts to support and integrate software and infrastructure supporting clinical trials should be consolidated under the NCI Division of Cancer Treatment and Diagnosis, with appropriate consulting and participation from CBIIT. This will ensure closer and more effective integration and coordination with cancer clinical trials efforts and needs. Efforts to support community-driven development and adoption of algorithms and software for basic research – such as the Knowledge Centers, the In Silico Research Centers, and open-sourced distribution of academia-based bioinformatics tools and algorithms – should be consolidated under the NCI Division of Cancer Biology. This will ensure optimal integration and cooperation with existing DCB programs, such as the Integrative Cancer Biology Program and Centers, and streamline the creation and usage of new computational and systems biology tools and databases for the cancer research community. Such separation and consolidation of informatics responsibilities with existing programs will ensure that software development and maintenance efforts will benefit first and foremost the cancer patients and the cancer research community.

**9. Use usual and established channels and mechanisms for concept clearance and peer review of NCI biomedical informatics initiatives in the future.** Funding for NCI biomedical informatics activities should be determined by established peer review criteria and should be supported by the documented needs of the cancer researcher community and the established track record of the investigators. Software tools produced intramurally by NCI researchers, as part of their lab activities, are not part of this recommendation and should be funded based on established principles and mechanisms for intramural research support.

**10. Promote interoperability and data sharing by making them key review criteria for grant and cooperative agreement applications and R&D contracts and including them as requirements for award.** Currently, there are virtually no NCI or NIH requirements for grantees to develop algorithms and tools that interoperate according to predefined standards. Similarly, while a data sharing section is required in all NIH grant applications, the terminology, activities, and enforcement of this section are essentially left undefined. The community would benefit greatly from more directed enforcement of data sharing requirements as well as from clear language that defines how and when data should be deposited.

## REFERENCES

1. Fearn, Paul, Recurring Themes in Informatics and IT from 60 Cancer Centers, National Library of Medicine Training Conference Denver, CO, June 15-16, 2010, and unpublished data
2. NCAB minutes, Dec 6-7, 1999.
3. NCAB minutes, June 12, 2002
4. NCAB minutes, June 10, 2003
5. NCAB minutes, November 2004
6. BSA minutes, March 2004
7. caBIG<sup>®</sup> Overview: Workspace and Working Group Kick-Off Meeting Handbook, February 2004 (provided by caBIG<sup>®</sup>).
8. caBIG<sup>®</sup> Strategic Plan 2005 (provided by caBIG<sup>®</sup>).
9. caBIG<sup>®</sup> Community Outreach Summit Executive Summary, January 2008 (provided by caBIG<sup>®</sup>).

## **Appendix A: List of WG Meeting and TC dates**

Meeting: November 2, 1010

Teleconferences:

November 22, 2010

December 1, 2010

December 8, 2010

December 15, 2010

December 20, 2010

December 27, 2010

January 5, 2011

January 12, 2011

January 19, 2011

January 26, 2011

February 2, 2011

February 9, 2011

February 16, 2011

February 18, 2011

February 23, 2011

February 25, 2011



## Appendix B: Agenda and List of Speakers - November 2, 2010

**NCI Board of Scientific Advisors**  
Ad Hoc Working Group on the  
NCI Cancer Biomedical Informatics Grid (caBIG<sup>®</sup>)

Marriott Bethesda North Hotel and Conference Center  
Bethesda, MD

November 2, 2010

### AGENDA

8:00 - 8:30 am	Call to Order and Introductions Discussion of Goals for the Day	Chair Members
8:30 – 8:55 am	Overview of caBIG <sup>®</sup> Program	Ken Buetow
8:55 – 9:15 am	Questions	Members
9:15 – 9:55 am	Clinical Science Efforts	John Speakman
9:55 – 10:15 am	Questions	Members
10:15 - 10:25 am	Break	
10:25 – 10:50 am	Executive Session	Members
10:50 – 11:35 am	Life Science Efforts	Julie Klemm
11:35 – 12:00 pm	Questions	Members
12:00 – 1:00 pm	Working Lunch/Executive Session	Members
1:00 – 1:35 pm	Research Infrastructure	George Komatsoulis
1:35 – 1:55 pm	Questions	Members
1:55 – 2:30 pm	Budget and Contract Management	Dwayne Forquer Mark Adams Greg Koreniewski
2:30 – 2:50 pm	Questions	Members
2:50 – 3:00 pm	Break	
3:00 – 3:20 pm	Executive Session	Members Ken Buetow
3:20 - 4:30 pm	Executive Session	Members

Mark Adams, Ph.D.  
Scientific Program Manager  
NCI cancer Biomedical Informatics Grid program (caBIG®)  
Principal, Booz Allen Hamilton  
McLean, VA

Kenneth Buetow, Ph.D.  
Director  
Center for Biomedical Informatics and Information Technology  
Chief, Laboratory of Population Genetics  
Center for Cancer Research  
National Cancer Institute  
Bethesda, MD

Dwayne D. Forquer, M.B.A.  
Chief of Staff  
Center for Biomedical Informatics and Information Technology  
National Cancer Institute  
Bethesda, MD

Juli Klemm, Ph.D.  
Associate Director  
Integrative Cancer Research Products and Program  
Center for Biomedical Informatics and Information Technology  
National Cancer Institute  
Bethesda, MD

George A. Komatsoulis, Ph.D.  
Deputy Director  
Center for Biomedical Informatics and Information Technology  
Acting Chief Information Officer  
National Cancer Institute  
Bethesda, MD

Gregory Korzeniewski, Ph.D.  
Director of Operations for caBIG®  
SAIC-Frederick, Inc.  
Frederick, MD

John Speakman  
Chief Program Officer  
Center for Biomedical Informatics and Information Technology  
National Cancer Institute  
Bethesda, MD

# Appendix C: Connectivity: Cancer Center Usage of caBIG<sup>®</sup> Capabilities (provided by caBIG<sup>®</sup>)

No.	Name of Center	Life Science	Clinical Trials	Infrastructure	No.	Name of Center	Life Science	Clinical Trials	Infrastructure
1	Baylor College of Medicine (Dan L. Duncan Cancer Center)	●	●	○	27	University of Alabama at Birmingham Comprehensive Cancer Center	●	●	●
2	Beckman Research Institute (City of Hope National Medical Center)	●	●	○	28	University of Arizona Cancer Center	●	●	○
3	Burnham Institute for Medical Research	●	○	○	29	University of California Irvine (Chao Family Comprehensive Cancer Center)	●	●	●
4	Case Western Reserve University, Ireland Cancer Center	●	●	○	30	University of California Los Angeles (Jonsson Comprehensive Cancer Center)	●	○	●
5	Dana Farber/Harvard Cancer Center	●	●	●	31	University of California San Diego (Rebecca and John Moores Cancer Center)	●	○	○
6	Dartmouth Medical School	●	●	●	32	University of California San Francisco Comprehensive Cancer Center & Cancer Research Institute	●	●	●
7	Duke Comprehensive Cancer Center	●	●	●	33	University of Chicago Cancer Research Center	●	●	○
8	Emory University (Winship Institute)	●	●	●	34	University of Colorado at Denver & Health Sciences Center (University of Colorado Cancer Center)	●	●	●
9	Fox Chase Cancer Center	●	●	●	35	University of Hawaii at Manoa (Cancer Research Center of Hawaii)	●	○	○
10	Georgetown University (Lombardi Comprehensive Cancer Center)	●	●	○	36	University of Iowa (Holden Comprehensive Cancer Center)	●	●	○
11	H. Lee Moffitt Cancer Center & Research Institute	●	●	○	37	University of Michigan Comprehensive Cancer Center at Ann Arbor	●	○	○
12	Indiana University (Melvin and Bren Simon Cancer Center)	●	●	●	38	University of Minnesota Cancer Center	●	●	○
13	Johns Hopkins University (Sidney Kimmel Comprehensive Cancer Center)	●	●	○	39	University of Nebraska Medical Center (Eppley Cancer Center)	●	○	●
14	Mayo Clinic Cancer Center	●	●	●	40	University of New Mexico Cancer Research & Treatment Center	●	●	○
15	Medical University of South Carolina	●	●	●	41	University of North Carolina Chapel Hill (Lineberger Comprehensive Cancer Center)	●	●	○
16	Northwestern University (Robert H. Lurie Comprehensive Cancer Center)	●	●	○	42	University of Southern California (USC) Norris Comprehensive Cancer Center	●	●	●
17	Ohio State University Comprehensive Cancer Center	●	●	●	43	University of Texas Health Science Center at San Antonio (San Antonio Cancer Institute)	●	●	●
18	Oregon Health & Science University Cancer Center	●	●	●	44	University of Texas M.D. Anderson Cancer Center	●	○	●
19	Robert Wood Johnson Medical School (Cancer Institute of New Jersey)	●	●	●	45	University of Virginia, Health Sciences Center (UVA Cancer Center)	●	●	●
20	Roswell Park Cancer Institute	●	●	●	46	University of Wisconsin Madison (Paul P. Carbone Comprehensive Cancer Center)	○	●	○
21	Salk Institute Cancer Center	●	○	○	47	Vanderbilt-Ingram Cancer Center	●	●	○
22	Stanford University	●	○	●	48	Virginia Commonwealth University (Massey Cancer Center)	●	●	●
23	Sloan Kettering Institute for Cancer Research (Memorial Sloan Kettering Cancer Center)	●	●	●	49	Wake Forest University Comprehensive Cancer Center	●	●	●
24	The Jackson Laboratory Cancer Center	●	○	●	50	Washington University St. Louis (Siteman Cancer Center)	●	●	○
25	The Wistar Institute	●	●	●	51	Wayne State University School of Medicine (The Barbara Ann Karmanos Cancer Institute)	●	●	○
26	Thomas Jefferson University (Kimmel Cancer Center)	●	●	●					

● Production Use     ○ No Activity  
● In Process     ● Planning

## Appendix D: List of Institutions Interviewed by Working Group Members

Amgen
Baylor College of Medicine/Dan L. Duncan Cancer Center
Case Western Reserve University
City of Hope National Medical Center/Beckman Research Institute
Dana Farber/Harvard Cancer Center
Dartmouth Medical School
Duke Comprehensive Cancer Center
Eastern Cooperative Oncology Group
European Bioinformatics Institute
ForteResearch
Fox Chase Cancer Center
Fred Hutchinson Cancer Research Center
Georgetown University/Lombardi Comprehensive Cancer Center
H. Lee Moffitt Cancer Center and Research Institute
Inova Health System
Johns Hopkins University/Sidney Kimmel Comprehensive Cancer Center
Memorial Sloan Kettering Cancer Center
Novartis
Ohio State University Cancer Center
Oregon Health & Science University Cancer Center
Roswell Park Cancer Institute
Salk Institute Cancer Center
Sanford/Burnham Institute for Medical Research
Stanford University
The Broad Institute of MIT
The Jackson Laboratory Cancer Center
The Wistar Institute
Thomas Jefferson University/Kimmel Cancer Center
University of Alabama at Birmingham Comprehensive Cancer Center
University of Arizona Cancer Center
University of California at San Diego/Rebecca and John Moores Cancer Center
University of California at San Francisco Comprehensive Cancer Center
University of California at Santa Cruz
University of Chicago Cancer Research Center
University of Iowa/Holden Comprehensive Cancer Center
University of Medicine and Dentistry of New Jersey
University of North Carolina at Chapel Hill/Lineberger Cancer Center
University of Pittsburgh Cancer Center
University of Southern California

University of Texas MD Anderson Cancer Center
University of Virginia Cancer Center
University of Wisconsin Madison/Paul P. Carbone Comprehensive Cancer Center
Velos
Wake Forest University Comprehensive Cancer Center
Wayne State University School of Medicine/The Barbara Ann Karmanos Cancer Center
Weill-Cornell Medical School

# Appendix E - caBIG<sup>®</sup> Workspace Structure (provided by caBIG<sup>®</sup>)

## Domain-level

Clinical Trials Management Systems Workspace (CTMS)

Integrative Cancer Research Workspace (ICR)

*In Vivo* Imaging Workspace (IMAG)

Tissue Banks & Pathology Tools Workspace (TBPT)

## Strategic-level

Strategic Planning Workspace (SP)

Training Workspace (D&T)

Data Sharing & Intellectual Capital Workspace (DSIC)

## Cross-cutting

caBIG<sup>®</sup> Vocabularies and Common Data Elements Workspace (VCDE)

caBIG<sup>®</sup> Architecture Workspace (ARCH)

## **Appendix F: List of caBIG<sup>®</sup> Service Providers (provided by caBIG<sup>®</sup>)**

5 am Solutions, Inc.  
Akaza Research  
Asclepius Solutions  
CTIS, Inc.  
E-SAC, Inc.  
Ekagra Software Technologies  
HealthCare IT, Inc. (HCIT)  
IMS, Inc.  
INFOTECH Soft, Inc.  
LabAnswer  
Moxie Informatics  
Persistent Systems  
Recombinant Data  
SAIC  
ScenPro, Inc.  
SemanticBits  
SRA Corporation  
TerpSys  
University of Utah

## Appendix G: Partial List of caBIG® tools in each category (provided by caBIG®)

### Integrative Cancer Research

GenePattern	Perform genomic data with powerful workflows
geWorkbench	Perform integrated research across genomic data
Bioconductor	Conduct high throughput genome analysis
caArray	Manage and annotate microarray gene expression data
caIntegrator	Develop custom, caBIG®-compatible web portals to conduct integrative research
Cancer Genome Workbench	Integrated Cancer Genomics Viewer
caBench-to-Bedside (caB2B)	Query caGrid for analytical and data services
caELMIR	Electronically manage laboratory data
caBIO	Obtain biomedical annotations from curated data sources
LabKey/CPAS	Analyze proteomics data
gridPIR	Query a database of genomic and proteomic annotations
Proteomics LIMS	Manage proteomics laboratory information
protExpress	Capture proteomics experimental annotations
Q5	Classify complex fragment mixtures in mass spectroscopy
QPACA	Analyze microarray data in the context of pathways
Reactome	Search a database of core pathways in human cancer
RProteomics	Analyze mass spectrometry proteomics data
SEED	Make and share genomic annotations
caGWAS	Analyze significant associations between genetic variations and disease
caMOD	Search a database of animal models for human cancer
Cancer Molecular Pages	Automatically annotate cancer related proteins
caNanoLab	Facilitate data sharing of nanoparticle information in cancer research
caTRIP	Query across caBIG® data services for clinical data
caCorrect	Correct artifacts and Improve the quality of collected microarray data
DWD	Distance Weighted Discrimination - Perform statistical corrections to reduce systematic biases in microarray data
caFE	Function Express - Analyze of microarray data using gene annotation data
GeneConnect	Map gene connections between different approved genomic identifiers
GOMiner™	Leverage the Gene Ontology (GO) to identify the biological processes
omniBiomarker	Identify differentially expressed biomarkers
Pathways Tools	Open source pathway database
TrAPSS	Screen and analyze RNA transcripts
VISDA	Visual Statistical Data Analyzer - Perform cluster modeling, visualization, and discovery

### Tissue Bank and Pathology Tools

caTissue Suite	Collect, store, annotate, aliquot, search, and track distribution of biospecimens; manage biorepositories
----------------	---



## Clinical Trials Management Suite

C3D (Oracle Clinical) C3PR	Collect clinical trials data Enroll, register, and track clinical trial participants across multiple sites
PSC	Create and manage clinical trial participant schedules and activities
Lab Viewer	Store, browse clinical laboratory data; share with other systems
caAERS	Collect and report adverse events
Integration Hub	Connect systems and support clinical trials workflow integration; provide interoperability associated with SOA
Clinical Connector	Interoperate / Share data with 3rd party CDMS systems
CTODS	Exchange of clinical trials data across multiple systems
FIREBIRD	Help investigators comply with Federal registration requirements
caMATCH	Recruit patients for clinical trials
CRF Project	Provide common data elements and case report forms for standardization and reusability

## In Vivo Imaging

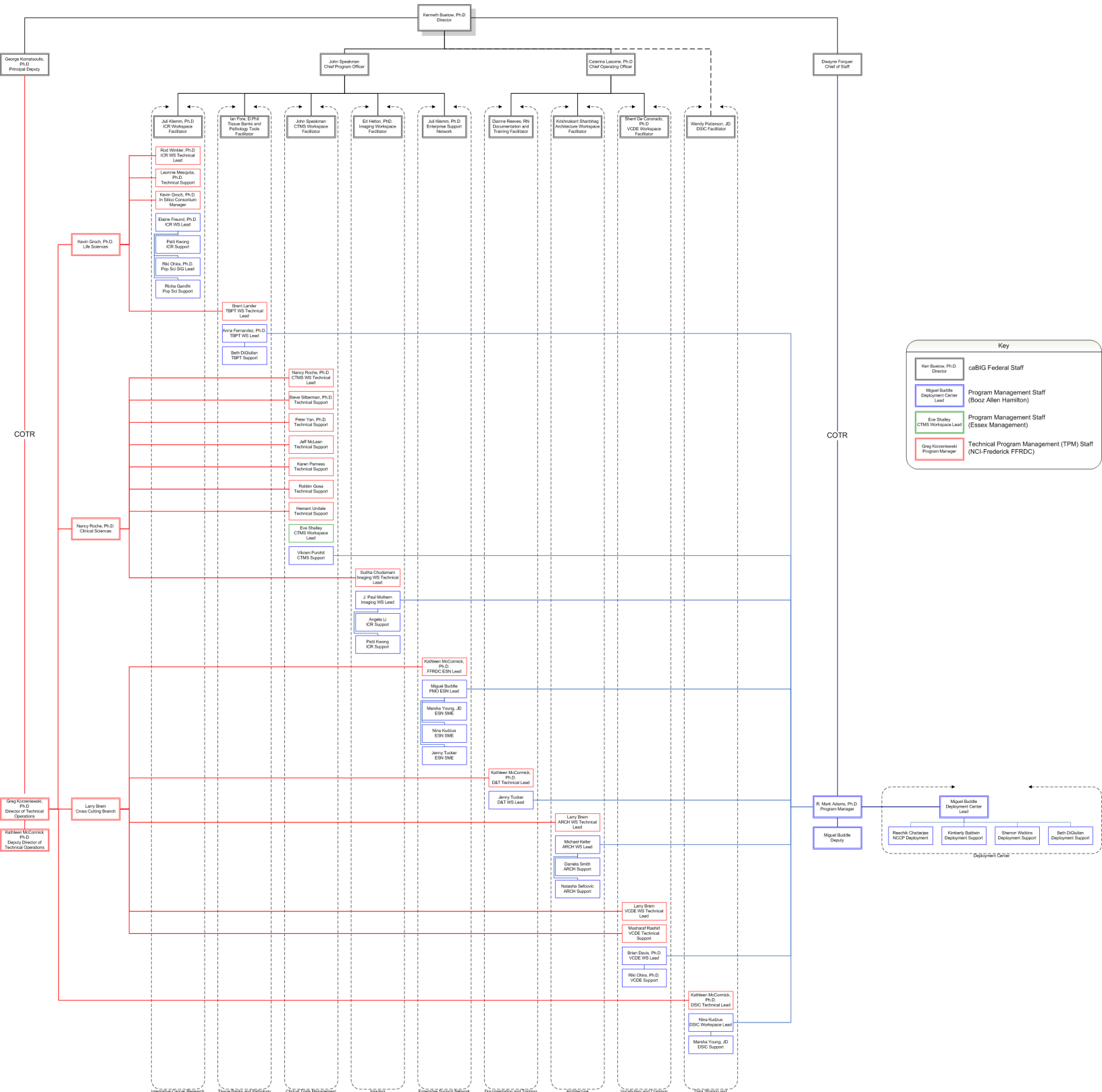
NBIA AIM	Store, annotate and share DICOM format medical images Annotation and Image Markup; Capture radiologist's notes and share with colleagues using standards-based annotations
caMicroscope XIP	Capture digital images of pathology slides eXtensible Imaging Platform Develop and rapidly test novel image analysis algorithms
Imaging Middleware/Virtual PACS	Connect commercial PACS systems with other 3rd party image databases

## Infrastructure

caGRID GAARDS	caBIG <sup>®</sup> interoperability infrastructure caBIG <sup>®</sup> federated security environment
------------------	---

# Appendix H: caBIG® Organizational Chart (provided by caBIG®)

## cancer Biomedical Informatics Grid (caBIG®) Organizational Structure



**Key**

- Ken Buntow, Ph.D. Director: caBIG Federal Staff
- Miguel Buitte Deployment Center Lead: Program Management Staff (Booz Allen Hamilton)
- Eve Shalley CTMS Workspace Lead: Program Management Staff (Essex Management)
- Greg Kozmowski Program Manager: Technical Program Management (TPM) Staff (NCI-Frederick FFRDC)

**Deployment Center**

- Miguel Buitte Deployment Center Lead
- Rachal Charney NCI/OP Deployment
- Himabey Baham Deployment Support
- Sharon Watkins Deployment Support
- Beth D'Gulian Deployment Support

Appendix I: caBIG® Budget Summary – Fiscal Years 2004-2010 (Appropriated, not including ARRA) (provided by caBIG®)

National Cancer Institute  
cancer Biomedical Informatics Grid® Program

Report to the Board of Scientific Advisors

Exhibit 1. caBIG® Budget Summary (FY2004 – FY2010)

	FY 2004	FY 2005	FY 2006	FY 2007	FY 2008	FY 2009	FY 2010
<b>Clinical Sciences</b>							
Clinical Trials Management System (CTMS)	\$1,607,791.06	\$1,607,791.06	\$5,785,400.66	\$5,791,893.65	\$13,038,731.20	\$15,580,069.99	\$8,309,309.94
In Vivo Imaging (IMAG)	\$0.00	\$0.00	\$1,706,301.04	\$4,279,931.84	\$2,151,090.41	\$3,209,903.34	\$1,901,440.65
<b>Life Sciences</b>							
Integrative Cancer Research (ICR)	\$3,855,597.51	\$3,855,597.51	\$5,289,737.79	\$5,819,519.58	\$8,951,761.51	\$6,268,267.20	\$8,599,831.11
Tissue Banks and Pathology Tools (TBPT)	\$2,220,832.75	\$2,220,832.75	\$3,050,990.42	\$1,550,887.90	\$2,048,881.44	\$1,318,601.63	\$950,000.00
In silico Cancer Research	\$0.00	\$0.00	\$0	\$0	\$0	\$3,690,196	\$3,691,960.87
<b>Cross-Cutting Workspaces</b>							
Architecture (ARCH)	\$1,081,681	\$1,081,681	\$3,508,652	\$4,183,516	\$7,976,320	\$4,990,629	\$7,268,516
Vocabularies and Common Data Elements (VCDE)	\$865,662.90	\$865,662.90	\$1,981,631.75	\$1,829,597.81	\$1,404,753.79	\$1,993,952.33	\$747,386.27
<b>Strategic Level Workspaces</b>							
Data Sharing and Intellectual Capital (DSIC)	\$534,580	\$534,580	\$708,833	\$333,923	\$280,408	\$182,527	\$89,214
Documentation and Training (D&T)	\$262,066.50	\$262,066.50	\$403,555.00	\$259,110.94	\$210,767.06	\$152,881.55	\$65,904.85
Strategic Planning	\$535,703.13	\$535,703.13	\$285,676.16	\$51,315.00	\$48,890.76	\$0.00	\$0.00
Knowledge Centers	\$0.00	\$0.00	\$0.00	\$0.00	\$5,944,377.00	\$5,391,535.00	\$4,612,708.00
<b>Contract management Cost</b>							
caBIG Total Program Management Cost	\$3,818,079.61	\$5,824,349.70	\$8,234,420.65	\$8,309,329.36	\$10,272,339.66	\$12,609,924.77	\$10,986,119.55
<b>caBIG Program</b>	<b>\$14,781,994.69</b>	<b>16,788,264.98</b>	<b>\$30,955,198.12</b>	<b>\$32,409,024.63</b>	<b>\$52,328,320.83</b>	<b>\$55,388,488.06</b>	<b>\$47,222,391.56</b>