

**Board of Scientific Advisors (BSA)  
caBIG® Oversight ad hoc Subcommittee Meeting**

Chicago O'Hare Airport Hilton Hotel  
July 25, 2011  
12:00 p.m. – 4:00 p.m.

**SUMMARY**

**Participants:**

Dr. Daniel Masys, Chair  
Dr. Brian Athey  
Dr. Andrea Califano  
Dr. Robert Comis  
Mr. Paul Fearn  
Dr. Gaddy Getz  
Dr. Joe Gray  
Dr. Rebecca Kush  
Dr. Lincoln Stein  
Dr. Lynn Vogel  
Dr. Jean Wang

**NCI Staff Participants:**

Dr. Harold Varmus  
Dr. Ken Buetow  
Mr. John Czajkowski  
Dr. Paulette Gray  
Ms. Claire Harris

**Introduction of Attendees—Dr. Daniel Masys and Subcommittee Members**

Dr. Daniel Masys asked the Subcommittee members and other participants to introduce themselves and describe their background as it relates to caBIG®.

**Charge From National Cancer Institute (NCI) Leadership and Expected Outcomes and Timeline—  
Dr. Harold Varmus**

Dr. Harold Varmus thanked the Subcommittee members and other participants for attending the meeting. He stated the purposes of the meeting: implementing the caBIG® Working Group Report recommendation to establish a scientific review board for caBIG®, setting guidelines for how the review board will operate, getting acquainted with other Subcommittee members and NCI Staff who will be working with the Subcommittee, and developing the missions of the Subcommittee. Dr. Varmus gave his perspective on the issues facing the Subcommittee. He recognized the importance of information technology and establishing standards for data exchange, which were the precepts on which caBIG® was founded. He stated that because of lack of oversight, both internal and external to the NCI, the program had become too large and expensive, and the level of use of its products was questioned.

A participant asked whether in the future caBIG® will strive to link with doctors that are not a part of the research community. Dr. Varmus responded that this is one of his main priorities.

**Review of the Findings of the caBIG® Working Group Report—Dr. Lincoln Stein**

A recurring theme of subcommittee discussions during this first meeting was the need to link real world scientific problems and researcher needs to CaBIG tool and resource development, as well as to ongoing activities such as standards development. The concept of “Driving Biological Projects” that underpin the National Centers for Biomedical Computing (NCBC) program of NIH is an appealing way to both anchor

the planning for data and applications functionality, and provide a yardstick for judging whether the subsequent output of a project is useful.

Dr. Lincoln Stein stated that the charge of the Working Group was to identify for the BSA the major challenges facing caBIG® and create a roadmap to make caBIG® more effective. He briefly outlined the Work Group's strategy, which included meetings with caBIG® leaders and contractors, interviews with users (both supporters and detractors), and internal teleconferences.

The principal problem with administration of caBIG® was perceived as an over-reliance on contractors without sufficient oversight by the NCI. Lifting the usual NCI controls for awarding subcontracts was supposed to make caBIG® more nimble, but led to caBIG® moving away from its original goals and the appearance of conflicts of interest in the contracting process.

Dr. Stein summarized the Work Group report's conclusions: caBIG®'s goals are critical for the NCI; the impact of projects it supported was not commensurate with the investment; the choice of what tools to develop was not driven by the users; it expanded too quickly; its approach was technology-centric rather than pragmatic; its software was not easily customizable and was expensive to support; and it lacked scientific oversight.

The Subcommittee added their comments on what they felt had been caBIG®'s major strengths and weaknesses. One participant noted that a key issue was that caBIG® had no critical driver projects. Dr. Masys suggested that caBIG® could learn from the success of the tools I2B2 and Research Electronic Data Capture (REDCap), developed by the Clinical and Translational Science Awards (CTSA) Consortium, which "went viral" because of their value and ability to accomplish a simple task well. Dr. Joe Gray felt that caBIG® projects lacked inclusiveness, being designed originally for NCI-designated Cancer Centers. Participants felt that one of caBIG®'s successes was establishing standards that were met its customers' needs. Dr. Rebecca Kush specified, however, that it was for establishing concept standards that caBIG received accolades. Dr. Masys noted that I2B2 and REDCap integrated standards after they were widely adopted, so that interoperability represented an enhancement rather than a barrier to adoption.

#### **Implementation Plan: caBIG® Working Group Report—Mr. John Czajkowski**

Mr. Czajkowski described caBIG® as being unevenly successful. He summarized the Work Group's recommendations, including redirecting caBIG® toward community-driven standards and stopping development of any projects outside its core mission. He lauded the report as being detailed, specific, clear, and easy to implement. The NCI's first response to its recommendations was to form a Scientific Advisory Group (SAG). He felt that establishing a dialogue between the NCI and the scientific community was the best path forward to reforming caBIG®.

He outlined the next steps to be taken in the caBIG® program. The NCI remains committed to the program, but would like to ensure that it is community-driven and meets the needs of the scientific community. The role of the SAG, comprised of the 11 members of this Subcommittee with one invitation pending, will be to help set priorities, a vital function with the program's significantly tighter budget. A participant suggested that a member of industry might be a good addition to the SAG.

Mr. Czajkowski outlined the qualities of the community-generated tools that will be supported by caBIG® in the future. They will be open source and will continue to be supported by the Support Service Providers (SSPs). Development will be guided by community needs and priorities. The NCI recognized

that the Knowledge Center program is critical and will ensure its continued support.

In developing new academic tools, caBIG® will collaborate with the NCI Scientific Divisions. Mr. Czajkowski observed that the scope, scale, and complexity of projects must be considered. He listed some possible projects, but added that it would be up to the SAG to decide what was funded and when. It is up to the SAG to determine how caBIG® will best fit into the system of systems.

Mr. Czajkowski said that he hoped this meeting would result in agreement about what caBIG® should do and what needs it will fill in the community. The NCI will partner with external, open-source developers who meet NCI standards to provide long-term support.

Mr. Czajkowski reviewed changes to the caBIG® budget. In fiscal year (FY) 2011/2012, the program's budget was cut to \$45 million dollars, compared to \$103 million dollars in FY 2009/2010. These budget cuts were a result of a decrease in appropriated funds by one-third and American Recovery and Reinvestment Act money by 50 percent. Phase-out costs make these budget cuts look less significant than in actuality. They represent a significant shift in funding. Mr. Czajkowski asserted that it will be SAG's task to determine which ongoing and new projects to fund given these significant cuts. He warned that the final budget for caBIG® depends on the budget for the NCI as a whole, which is still uncertain.

A participant asked how many of the budget cuts would come from funding for contractors.

Dr. Ken Buetow responded that all cuts would be in the contracting budget with none from academia.

The question was asked how the relationship between the NCI and Booz Allen and Science Application International Corporation (SAIC) will change. Mr. Czajkowski replied that this would be covered later.

The Subcommittee discussed what they felt would be the central issues in the future for caBIG®.

Dr. Andrea Califano suggested that it was important to make these sophisticated tools accessible to the broader community. When asked about interoperability between the research and clinical communities, Dr. Califano responded that attempts to unify clinical informatics and bioinformatics departments had failed. He felt that recognizing the fundamental differences between the two groups would be the best way to meet the needs of both and create a bridge between them. Dr. Masys volunteered that there were actually three communities: molecular genetics researchers, clinical researchers, and clinical practitioners. A participant argued that tool development should be informed by the user communities' needs and genetic testing information was of little practical use to a clinician. Another countered that science changes and it was important that tools be built for changing needs. Another participant added that an important type of tool that does not exist is translators that would allow for changes in vocabulary over time.

Mr. Czajkowski concluded that he thought that a set of criteria was emerging from the discussion and suggested that it would be worthwhile to publish it.

#### **caBIG® Program Goals Overview And Program Accomplishments—Dr. Ken Buetow**

caBIG® pilot phase goals, focused exclusively on NCI-designated Cancer Centers, were to have a common, widely distributed infrastructure that permitted the cancer research community to focus on innovation; a shared vocabulary, and common data elements and data models to facilitate information exchange; a collection of interoperable applications developed to a common standard; and raw published cancer research data available for mining and integration.

The original strategy was to create an open community of participants, adopt a federated model, support local activities, and allow groups to share either with their collaborators or in a framework that can be

generalized. Many of the groups have adopted within-organization federation, not necessarily national or larger scale federation. The goal was to make sure that the information and data being generated in the cancer enterprise were accessible. caBIG® leveraged existing academic efforts and commercial efforts to invest primarily where there was a need for new capabilities.

As of 2010, there were 56 NCI-designated Cancer Centers utilizing enterprise-wide adoption of many components, along with 30 NCI-designated community cancer center programs and six in silico research centers. The program has numerous vocabularies, data element information models, and specifications. There were a wide range of applications developed to respond to specific community requests, but many were retired quickly as they did not achieve penetration. In addition, caBIG® has created seven knowledge centers that facilitate the use of biomedical informatics tools organized around particular domains. The 19 licensed SSPs are essentially commercial entities whose business model is to support caBIG activities. The SSPs are places where service-level agreements can be created to maintain support for particular types of activities.

A participant asked if the NCI ever created criteria for the meaningful use of the tool for science. Dr. Buetow responded that listed entities were using some part of caBIG® to support a specific need. Another participant commented that people were labeled as users if they had downloaded the software, and that there was a distinction made within the last 2 to 3 years distinguishing between adopters and adapters.

Dr. Jean Wang stated that the NCI has given all Cancer Center Support Grants (CCSGs) supplemental funding to adopt caBIG®. Some smaller Cancer Centers are actually heavily dependent on caBIG® for their informatics support.

A participant noted that his Cancer Center just went through recompetition and had to implement some part of caBIG® to apply; if people are paid to do something, they may do it even whether they want to or not. Dr. Buetow stated that there are more than 70 tools in caBIG® and some, such as gene patterns, are very widely used; if gene patterns are counted as an adoption, there will be a very populated map. Dr. Masys commented that those are things that show value and researchers do not use them for their caBIG® label. Another participant asked to see the data on broad usage because the Subcommittee needs clear evidence of its value. Dr. Buetow stated that he would share the data, and that caBIG® made substantive investments in programs like gene patterns. Dr. Wang agreed that gene patterns are a great example of how caBIG® added value.

A participant stated that the goal of this oversight subcommittee is to accelerate the process by which caBIG® identifies potential winners and also accelerate the process by which the failures are identified and eliminated. Another participant added that caBIG® should not coerce behaviors that are not natural to the research and productivity of the Cancer Centers.

A participant stated that it is necessary to articulate what success means for caBIG® or the NCI will not be able to demonstrate any meaningful progress on the project. Dr. Stein said that it was difficult to measure whether the caBIG® infrastructure actually catalyzed science. It moved researchers forward in observable and measureable ways and that would not have occurred if caBIG® had not had its input, but those are very difficult outputs to quantify and measure.

Dr. Gray suggested that a better way to show caBIG®'s impact on the community than just dots on a map was needed. Dr. Buetow agreed that meaningful measures were needed. A participant asked how often code was contributed by people who are not part of the program itself.

Dr. Buetow responded that there had been a lot of input and contribution of code from places where the source code has been deployed. One of the strategic directions that caBIG® wants to take is to leverage true open-source community activities.

In response to the BSA report, caBIG® maintained academic-led efforts, capabilities identified as having success within the community, biospecimen management, imaging, multi-institutional clinical trials and the components necessary to support them, standards, and interoperability efforts; changed community engagement to be through NCI staff instead of contractors; discontinued software efforts through contractors; updated architecture; worked to create electronic health record specifications and reference implementations; and suspended new initiatives until there is community input. A participant asked if there was a list of the specific programs that have been maintained. Dr. Buetow responded that he could provide such a list.

Dr. Wang asked if the existing software still was being supported. Dr. Buetow replied that everything caBIG® had developed still is available for download with all of its corresponding materials, but the vast majority of software is not under active maintenance. A participant asked if the public was aware that improvement of tools has now ceased. Dr. Buetow answered that most people are aware of what is ongoing and what has stopped.

A participant noted that the subcommittee ought to dictate that a new software product must have certain functional characteristics that contribute to the overall goals of caBIG® to be useful.

Dr. Buetow commented that it was not caBIG®'s intent to develop software de novo, and the core development has been driven by academic colleagues.

Dr. Gray held that in terms of a long-term sustainability plan for caBIG®, what has been missing is a handoff to the private sector. Dr. Buetow answered that he hoped that the original academic investments can be brought into the commercial world. Dr. Gray added that this message had not been communicated to the private sector because several software development companies had said that they did not invest in this area because it was caBIG®'s purview.

A participant noted that incentivization creates organizational behaviors that muddle the value of caBIG®. Dr. Wang added that small start-up companies perhaps should be incentivized. Dr. Buetow noted that incentives were for caBIG® compatibility. Any vendor who had standard compliant products could and did compete.

In terms of standards and interoperability connectivity, the caBIG® community will continue to broker pre-competitive specifications used to drive creation of software by academic and commercial developers.

A participant asked for more specific priority areas. Dr. Buetow responded that those listed were broad enough for caBIG® and the Subcommittee to raise ideas that might have broad traction across the communities. A participant suggested that one of the ways to do this is to request a white paper from the community that would discuss the significance of this development driving biological problems and potential approaches.

A participant noted that a big question for caBIG® was if it is going to fund discovery in informatics or if it is a tool-building effort to help scientists get their job done. This has been a source of tension in the clinical and translational science awards informatics cores as well. Some of them want to use it as if it is an R01 just to advance knowledge representation and natural language processing.

Dr. Masys said that he had been writing down features of the functional specifications that would make something eligible for caBIG® support if the Subcommittee can articulate them so that it becomes good guidance for the program. He will draft the first version and see if members agree on what the Subcommittee's input should be to the NCI. He will send it to the Executive Secretary to distribute to the members.

### **Members' Requests for Additional Information**

To facilitate the review and prioritization of caBIG® tools, the subcommittee believes that a useful way to organize caBIG® resources and software applications for tracking and review would be to put them in one of four categories:

- 1) Resources and applications that are actively being enhanced to expand their functionality to meet user demand, for which timelines and future deliverables are specified;
- 2) Resources and applications that are being actively supported but for which there is no specific plan for future enhancements, no driving scientific problem, or very limited user communities;
- 3) Previously developed resources and applications that are no longer being supported but are available for download from the caBIG® archive, and:
- 4) "New opportunities" representing program expansion in novel areas.

The subcommittee requests that a detailed tabular summary of all caBIG®-supported projects and activities be created and updated periodically as a synoptic management tool that the subcommittee can use to assess program directions.

The subcommittee also requested that the current usage statistics of caBIG®resources be supplemented by measures that represent 'meaningful use' of the applications rather than simply numbers of downloads or numbers of executables installed and running successfully. Ideal measures would include some form of measurable impact on scientific productivity as a result of their use. One possible approach in this area that has been implemented by CTSA awardee institutions is to do automated literature surveillance of newly published articles by faculty at the institution, and send an email with a hyperlink to a brief online survey tool that asks the researcher whether the work described in the publication employed CTSA-supported resources, and if so, which ones. Notably, responding to this automated message is a condition of continued CTSA support at some institutions. Other more traditional process (rather than outcome) measures would include numbers of projects/investigators supported. Having such user-reported 'meaningful use' data will enable the subcommittee to more accurately assess the uptake and utility of caBIG®-supported tools and resources, as seen by their intended users.

The subcommittee was also interested in further information regarding specific programs in caBIG®, including the activities and involvement with the caBIG® Support Service Providers program, information on specific projects utilizing caBIG® such as I-SPY and In Silico Research Centers, and information about successful community driven projects such as REDCap and i2b2.

### **Subcommittee Work Plan**

In its formative discussions regarding how it would do its work of assessing caBIG® program effectiveness, the subcommittee empirically created a set of functional criteria it would use to judge new and existing activities. Some or all of these functional criteria may be useful to include in the future as elements in competitive solicitations issued by caBIG® program staff for new projects:

- 1) Does the activity, application or resource meet a well-articulated and attainable need of basic, translational or clinical researchers or cancer health care (ie., is there a 'driving biological or clinical project' and are the intended users members of the project team)?
- 2) How will success or failure be evaluated? Analogous to stopping rules for clinical protocols, what will be the stopping rules for ending the project if it either fails to meet its technical objectives or fails to be adopted even if technically successful?
- 3) Will the activity, resource, or application, if successful, make some objectively measurable incremental progress toward the overall caBIG® vision of interoperability of data and systems? Will it enable data sharing and make use of and/or enhance open international standards for research? Will it follow the development principles of caBIG®?
- 4) Is the activity, resource or application designed to anticipate change in a rapidly expanding knowledge base of science and practice? Flexibility and generalizability are important characteristics for longevity in an era of agile science that is changing at a high rate.
- 5) Is the intended deliverable of the project achievable in the time frame and budget proposed?
- 6) Will the output of the project be broadly implementable by organizations of varying size and sophistication? Will it be used broadly by organizations and institutions outside of NCI/Cancer Centers (e.g. other NIH centers or academic research organizations)?
- 7) Is there a documented plan for long term maintenance, enhancement and fiscal sustainability of the activity, application or resource and its user base? What is the user base and has there been a stakeholder assessment to assure that the activity, application or resource will indeed meet a currently unmet need?
- 8) Is the project generalizable and likely to create value or address broad needs across the community of cancer centers and investigators? Or would this activity, resource or application be perceived as a "pet project" of an "in" group?
- 9) Does the activity, resource or application have enough market value to gain adoption without incentives, or if financial or policy incentives are required, are they justified?

With respect to providing support for applications and resources developed by academic institutions, the subcommittee notes that there are several successful models already in use at NIH. These include small, peer-reviewed "seed funding" grants or contracts for exploratory development of prototype tools or systems, and support for scale up and enhancement of already-developed systems and resources. The established 'software enhancement' grants programs supported by a number of Institutes require that the system developers document current and proposed system functionality, intended deliverables, and provide documentation of an existing user base, and input from users that the proposed software enhancements are needed and will be used and will follow caBIG® principles in terms of software development.

The original caBIG® priorities were based on challenges articulated by the NCAB circa 2002, and a subsequent survey of cancer centers in 2002. The informatics landscape has changed tremendously since that time, so a new systematic survey of needs may be warranted.

With the additional information noted above, the subcommittee anticipates creating an effective and ongoing oversight interaction with caBIG® program staff and NCI leadership. As an initial workplan, the committee will initiate a monthly conference call to review progress in implementing the caBIG® operations model recommended in the March 2011 report, to review the detailed list of caBIG® projects and discuss new programs and projects whose review is requested by NCI CBIIT, and any other items whose review is requested by NCI leadership.

Dr. Varmus noted that he liked the idea of soliciting white papers on broad topics and using these as a

good vetting area for intentional requests for applications (RFAs).

### **Additional Information**

Dr. Masys noted that the Subcommittee should identify specific, deep community needs. There is not a need to set future meeting dates if members are comfortable with group editing of documents and with a monthly call. The Subcommittee should hold a call once per month between now and roughly 6 months from now. Reporting to the BSA as a whole would be useful.

**Adjournment – Dr. Daniel R. Masys**

There being no further business, the meeting was adjourned at 4:10 p.m. on Monday, 25 July 2011.