



## The Cancer Biomedical Informatics Grid (caBIG): Pioneering an Expansive Network of Information and Tools for Collaborative Cancer Research

Kerry K. Kakazu PhD, Leo W-K Cheung PhD,  
and Lynne Wilkens DrPH

The Cancer Research Center of Hawaii was invited in the summer of 2003 to participate in the cancer Biomedical Informatics Grid (caBIG). The project was initiated by the National Cancer Institute's Center for Bioinformatics (NCICB). Its ambitious goal is to build a network for cancer-related biomedical informatics that will ultimately enable information to be shared easily between cancer researchers. The NCICB believes that successful completion of the project will serve as a model for how all biomedical research can be streamlined to serve the needs of patients, study participants, care providers and researchers more effectively.

### The Informatics Challenge:

Technological advances in biological data collection have resulted in a tremendous surge in the amount of data to be stored and processed. The study of informatics involves the application of statistical methods along with computer and information sciences to organize, manipulate, analyze and visualize these large data sets. That task in itself is daunting, but an additional layer of complexity results when attempting to integrate data stores between different disciplines and researchers. The issues facing the management of clinical trials participants are different from those involved in the processing of data generated from a DNA microarray or the analysis of questionnaires for a population study. Even within many disciplines, there are no standards for data collection, making integration of data among researchers difficult. However, if cancer researchers are to gain the maximum utility

from the data they are collecting, linking data from different studies and across disciplines will be necessary. Providing the foundation and tools to accomplish this task is the challenge that the NCICB has undertaken.

### The caBIG Proposal:

The vision of the caBIG project is to have the NCICB coordinate the development of a unifying informatics architecture. Key to the project is the utilization of the NCI Cancer Centers framework to identify the resources and expertise in cancer informatics that exists among its facilities and researchers. The combined resources of the Cancer Center participants and NCICB can then be used to provide the linkages enabling data sharing across disparate databases. Ultimately, standard vocabularies, data structures and analyses tools will be developed that can tie together data generated from basic

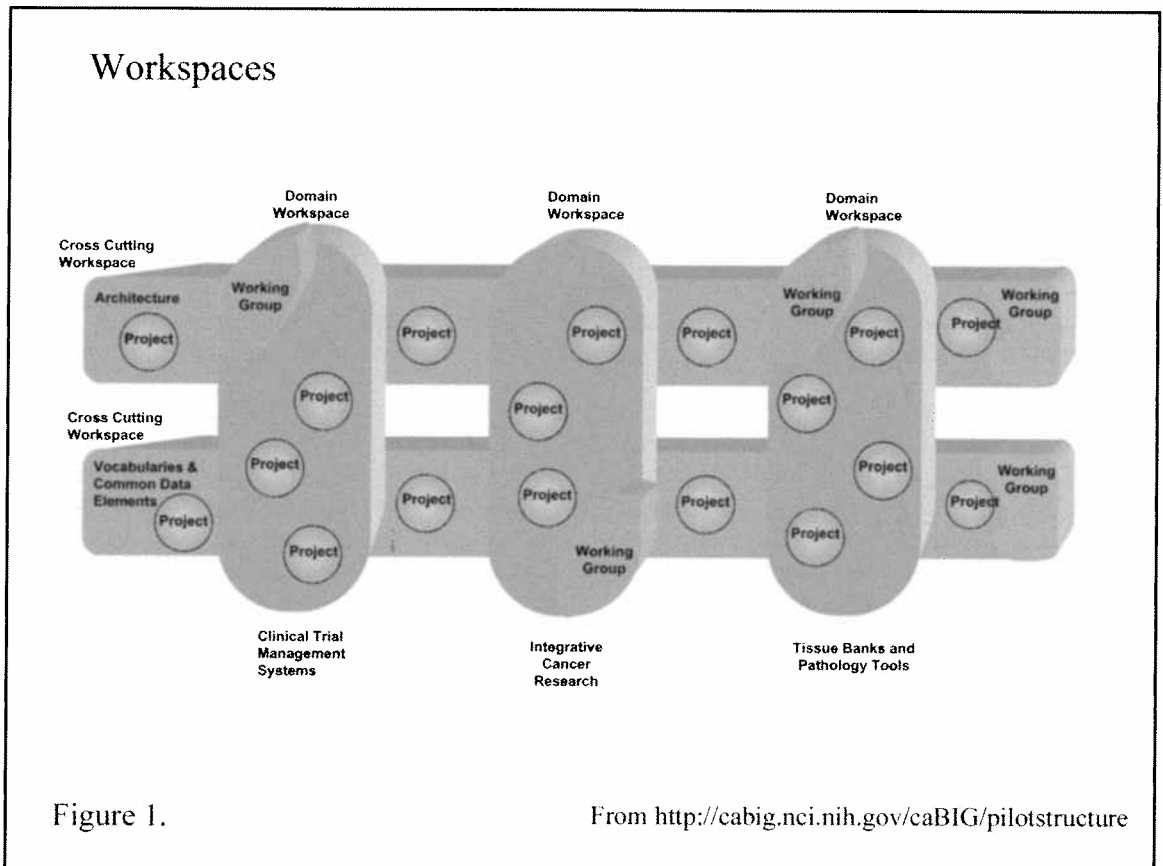


Figure 1.

From <http://cabig.nci.nih.gov/caBIG/pilotstructure>

science and clinical research. The increased integration should lead to the discovery of data relationships that might otherwise be overlooked.

### **Pilot Phase Development:**

The Cancer Research Center of Hawaii (CRCH), as one of the 61 NCI-designated Cancer Centers, was asked to become a part of the pilot phase of the caBIG. The original structure of the pilot phase proposed the selection of approximately 10 Cancer Centers to work on informatics projects that would reflect each center's research strengths, show the feasibility of developing an informatics grid and generate tools to be used by other researchers. CRCH submitted a proposal that focused on data integration issues in the study of the lifestyle and genetic factors contributing to the ethnic variations in cancer incidence in Hawaii.

After evaluation of all the proposals and meetings between NCI and Cancer Center representatives, the NCICB decided to expand the scope of the pilot phase by including more centers and grouping similar projects with a common focus into informatics "Workspaces". The five Workspaces proposed are divided between "Domain Workspaces" that will develop products to address an informatics need identified by the various cancer centers, and "Cross Cutting Workspaces" that will address issues affecting the Domain Workspaces such as infrastructure, vocabularies, common data elements, standard documentation and data security. Within each Workspace a center may have a role as a developer or an adopter depending on its capabilities and needs (Figure 1).

#### ***Domain Workspaces***

**Clinical Trials Management** – The management of clinical trials participants was the most frequently cited data management challenge for all Cancer Centers. A variety of software tools designed to aid clinical trials data management is already available. This Workspace will work to identify and assemble those tools that fit into the caBIG framework. There will be an emphasis on the utilization of "open source" products to promote universal access to the tools. Cancer Center members of this Workspace will help to develop new tools to add additional capabilities to the software grid.

**Integrative Cancer Research** – This Workspace is focused on enabling data integration and sharing between different researchers. Its objective is to provide tools that will link data sets between different researchers as well as between different disciplines of research. The Workspace aims to demonstrate the feasibility of creating a shared informatics platform that will be available to the entire cancer research community. The diversity of issues to be addressed in this Workspace has resulted in the formation of a number of sub-groups, or special interest groups (SIGS) to work on issues of specific concern to different areas of study. As an example there are SIGs for discussion of DNA microarray data repositories, data analysis and statistics, and translation of basic research to clinical treatments.

**Tissue Banks and Pathology Tools** – The primary goal of this Workspace is the development of tools to inventory, track, mine and visualize tissue samples available from a variety of sources. As is the case with clinical trials data management, a number of tools have been developed independently, but this Workspace will assist

in the unification of the resources available. A unique aspect of this area of interest is the storage of physical samples, thus issues of geography are a key consideration in the creation of effective tools and in linking the existing databases describing the samples.

#### ***Cross Cutting Workspaces***

**Vocabularies and Common Data Elements** – The purpose of this Workspace is to address software development issues relating to the creation of tools that will fit the caBIG information network model. Its members will work to establish standards for the tools used in caBIG. They will also be responsible for assessing existing systems that may be integrated into the project.

**Architecture** – This Workspace will put its emphasis on developing the computing framework to support caBIG. Its topics include methodologies for data access and presentation, data transmission standards and data security issues. It is also tasked with ensuring consistent application of development principles of the caBIG.

#### **Role of CRCH**

CRCH has been selected to participate as a developer in the Vocabularies and Common Data Elements (VCDE) Workspace, one of the Cross Cutting Workspaces. As described above, CRCH will contribute to the implementation of software development standards.

Two researchers in the Biostatistics group of the CRCH Cancer Etiology Program, Dr. Leo Wang-Kit Cheung and Dr. Lynne Wilkens are our representatives in this Workspace. Dr. Cheung has a Ph.D. in Statistics with a background in computer science, genetics and molecular biology. His research interest in bioinformatics, especially as related to genomic and proteomic data mining and analyses, reflects his varied experience. Dr. Wilkens has a Doctorate of Public Health in Biostatistics and heads the Biostatistics Shared Resource at CRCH. Her research focuses on techniques for studying disease associations when the independent variables are measured with error. CRCH's main contribution to the VCDE Workspace will be to add functionality in terms of vocabulary and common data elements for cancer epidemiology, particularly nutritional epidemiology. The Cancer Etiology Program of the CRCH has extensive experience in studying the role of diet in cancer incidence and will aid in the publication of select international vocabulary systems for dietary variables, as well as CRCH's internal system, on the caBig. These systems will aid researchers in the future to collect dietary data in a uniform fashion, allowing for easier data integration. Drs. Cheung and Wilkens will also serve as VCDE liaisons to the Integrative Cancer Research Workspace.

The caBIG project and the amount of resources that the NCI is contributing to its operation demonstrate the importance that informatics now plays in biomedical research and treatment. The Cancer Research Center of Hawaii is pursuing other partnerships that will help to develop our resources in this area. The CRCH has entered into collaborations with the Maui High Performance Computing Center (MHPCC), an Air Force Research Laboratory Center managed by the University of Hawaii. Dr. Cheung will utilize the capabilities of the supercomputer facility to support his research in bioinformatics. The MHPCC and the CRCH Etiology program

will work together to develop a secure web-based data management system for a large epidemiology project that will allow easy access to and integration of data for researchers. Dr. Cheung will also participate in the University's Center for Genomic, Proteomic, and Bioinformatic Research (CGPBR), a new initiative that hopes to aggregate and support investigators working in these areas.

### Conclusion

The caBIG is a major undertaking to direct cancer informatics toward a unified architecture that facilitates data sharing and integration. The resulting increase in data mining capabilities as well as expanded opportunities for collaborative research should ultimately lead to faster and more effective development of potential treatments and prevention programs. The commitment of the NCI to the success of this project, along with the challenges CRCH faces in managing and integrating its own data stores from epidemiological studies, genomics research and clinical trials, makes our participation vital to the continued development of our research capabilities.

For more information about the Cancer Research Center of Hawaii, please visit our website at [www.crch.org](http://www.crch.org).

### References

1. National Cancer Institute. cancer Biomedical Informatics Grid (caBIG). Available at: <http://cabig.nci.nih.gov>. Accessed July 26, 2004.

## Pregnant and professional – we've got you covered.

As Oahu's only source for better maternity and nursing wear, we proudly provide maternity wardrobe for Trini Kaopuiki of KHON and the Baraquios of "Living Local" on Oceanic 16.

University Square, Suite A308 (3rd Floor)  
Corner of South King St. and University Ave.

makana  
mother  
& baby

942-baby

Friday & Saturday • 11 am – 6 pm  
Monday – Thursday • by appointment  
[www.makanamotherandbaby.com](http://www.makanamotherandbaby.com)

# Are you paying a hefty price for your medical malpractice coverage?

To protect your practice, you must have medical malpractice coverage. But at what cost? Medical malpractice premiums are on the rise, nationally. Some carriers have had substantial rate increases. Other carriers have pulled out of the market, leaving their insured searching for alternative coverage. Is there anyone you can count on in this time of uncertainty?

### **You can count on HAPI.**

HAPI has been a reliable Plan for Hawaii's physicians for 25 years.



Started 25 years ago, HAPI is Hawaii's first, physician-owned, medical malpractice coverage provider. To learn more about HAPI and the cost savings it may offer you, call Jovanka Ijacic, our Membership Specialist.

### **HAPI's Physicians' Indemnity Plan**

735 Bishop Street, Suite 311, Honolulu, HI 96813  
Ph: 538-1908, [www.hapihawaii.com](http://www.hapihawaii.com)