National Cancer Institute

EMPOWERING COLLABORATION

Across the Cancer Community

caBIG® ANNUAL REPORT 2009

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

caBIG® 2009 At a Glance...

caBIG® program summary:

- 2,300+ people participating in the program from more than 700 institutions
- 1,050 attendees at the 2009 caBIG® Annual Meeting, more than 600 of whom were there for the first time
- 70+ organizations connected to caGrid, the interoperable IT network designed to facilitate data sharing, including 56 NCI-designated Cancer Centers and members of the NCI Community Cancer Center Program (NCCCP)
- 30 peer-reviewed scientific publications in which caBIG® tools and technology enabled the research or were the main focus of the paper
- 1.19 million biospecimens available through caGrid
- 3.71+ million medical images stored in the National Biomedical Imaging Archive (NBIA)
- 25,000+ microarray experiments available for research use on caGrid
- 15 licensed Support Service Providers to sustain the biomedical community as they deploy caBIG® tools and technology
- 15 countries using or evaluating caBIG® tools and technology to facilitate collaborative biomedical research, in cancer and beyond

caBIG® Vision

A virtual network of interconnected data, individuals, and organizations that redefines how research is conducted, care is provided, and patients/participants interact with the biomedical research enterprise.

caBIG® Goals

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

CONTENTS

2 LETTER FROM THE DIRECTOR, CENTER FOR BIOMEDICAL INFORMATICS AND INFORMATION TECHNOLOGY

3 WHAT IS A 21ST CENTURY BIOMEDICAL NETWORK?

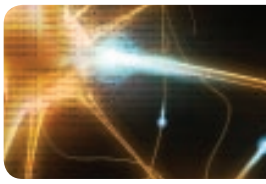
5 caBIG® LINKS THE CANCER COMMUNITY

9 caBIG® IN ACTION

18 NEW MODELS OF COLLABORATIVE CANCER RESEARCH AND CARE

24 OUR ROLE IN THE FUTURE OF BIOMEDICINE

3

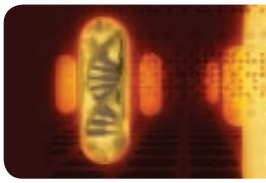


5

9

18

24



TO FRIENDS AND COLLEAGUES THROUGHOUT THE BIOMEDICAL COMMUNITY:

The year 2009 (to paraphrase Winston Churchill) could be described as the "end of the beginning" for the caBIG® initiative.

Our first six years have been devoted to laying the groundwork for a new era in which data interoperability links the entire cancer community nationally and globally. Specifically, our initial charge was to enhance the ability of NCI Cancer Centers to leverage biomedical data for research; connect workflows and streamline data collection; increase accuracy of processes; perform complex analysis across data sets; identify best practices; and share data for collaborative research. In other words, our mission has been to help each Cancer Center to address its complex research and care challenges.

As a result of the dedicated efforts of our community of thousands of individuals from more than 700 institutions, we are proud that today:

- caBIG® provides a rich collection of standards-based infrastructure and tools to facilitate and accelerate the critical functions in R&D processes.
- caBIG® technology lets any organization break down silos within its four walls, as well as overcome information barriers between each organization and the outside world.
- caBIG® web-based services enable organizations to partake of capabilities in "a-la-carte" modules or comprehensively across the discovery-development-delivery continuum to retain maximum flexibility and cost-effectiveness.
- caBIG® provides the capabilities to ensure access and use of data from major science initiatives such as The Cancer Genome Atlas.
- caBIG® usage keeps getting easier, facilitated by a diverse collection of support services provided by government, academic, and commercial sources.

Our next challenge, however, has just begun. Biomedicine is still decades behind in the "knowledge economy" curve. Translational research and personalized medicine require integration of multiple modalities and dimensions of data, and such integration has largely not yet been achieved in the cancer community. As a result, data are still locked away in incompatible formats and systems. Research studies are still slow and designed "one at a time": clinical trials take too long to initiate, too long to accrue patients, and far too long to report outcomes that could fuel new approaches.

We are tackling these challenges in the next phase of our caBIG® initiative.

First and foremost, we are striving to meet the evolving needs of the cancer community. We are eliciting input from senior leaders, basic and clinical researchers, and informatics professionals about what "success" means at their institutions.

Second, we are continuing our collaboration with the American Society of Clinical Oncology (ASCO) to define a specification for an oncology-extended Electronic Health Record, thereby ensuring that as the huge U.S. commitment to digitalization of medicine unfolds, oncology can seamlessly interoperate among clinical care, clinical research, and regulatory reviews. In addition, by helping in this Health IT transformation, caBIG® capabilities provide a leg up in comparative effectiveness research and the development of a strong evidence base achievable through the large-scale collection and analysis of clinical outcomes.

Third, we are exploring new models of research that exploit the Internet and our world of instant communications and ubiquitous connectivity, empowering consumers and patients to play a central role in research.

Fourth, we are strengthening the ecosystem that caBIG® has pioneered through the BIG Health Consortium™—comprised of all biomedical sectors—via strategic partnerships to conduct projects that demonstrate genuine 21st century biomedicine.

Fifth, we are facilitating the use of caBIG® infrastructure in domains beyond cancer, to speed the benefits of interoperability to research and care in all diseases.

Cancer is a complex and daunting disease. But the greater cancer community—linked through caBIG® technology to make optimum use of a "world of data"—is resolute in striving to accelerate research discoveries and improve patient outcomes.

Ken Buetow, Ph.D.
Director, Center for Biomedical Informatics and Information Technology
National Cancer Institute
April 16, 2010

What is a 21st Century Biomedical Network?

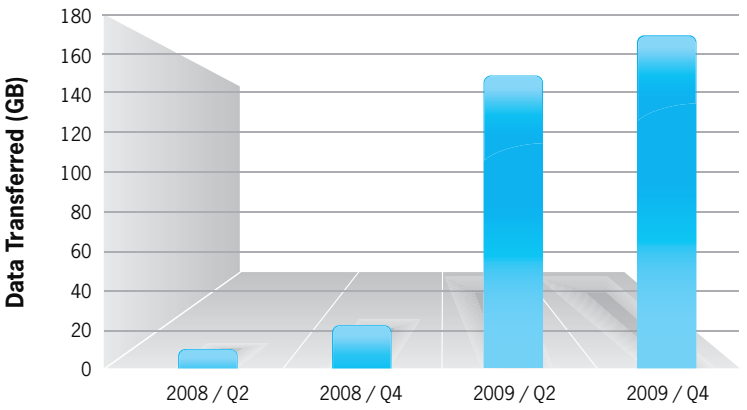
We know that to achieve the promise of a new era in biomedicine—and to deliver care that is personalized, predictive, preemptive, and patient-centered—we must capture and analyze huge amounts of biological and clinical data, disseminate it widely, and continually transform it into knowledge.

Thus, the concept of a 21st century biomedical network encompasses several requisites:

- A flexible, standards-based information technology infrastructure that facilitates rapid flow of information with appropriate security and privacy protections;
- A community of stakeholders who share a common desire to improve clinical care and accelerate research discoveries by generating, using, and sharing data; and
- An expanding collection of data resources.

READ ON OR VISIT US ONLINE AT: CABIG.CANCER.GOV/21STCENTURY

Increased Data Sharing



The volume of gene expression data downloaded via the caBIG® tool known as caArray has increased more than 8-fold in less than 2 years.

> "It [caBIG®] literally, I believe, is as easy as buying a computer and getting wireless to get in the game. What's so exciting about this is that you don't have to have millions of dollars ....to mine the data."

Michael Caligiuri, M.D.
Director of the Comprehensive Cancer Center and CEO of the James Cancer Hospital and Solove Research Institute, The Ohio State University

To address this challenge, the National Cancer Institute developed the **caBIG®** (**ca**ncer **B**iomedical **I**nformatics **G**rid®) initiative. caBIG®—through a portfolio of tools, grid infrastructure, vocabularies, policies, working groups, community events, publications, websites, data portals, training, Knowledge Centers, service providers, and countless other resources—has enabled a "national network of cancer research and care". Through caBIG®, we are:

Turning Growing Data Volumes from Liability to Asset—Biomedical research generates terabytes of data. caGrid (the caBIG® grid infrastructure) provides proven standards-based technology for organizations to securely store, analyze, and share all types of data, including medical images, genomic information, biospecimen annotations, or clinical information.

Making Data "Liquid"—Data from diverse areas of research must be interoperable so they can be analyzed in an integrated fashion, and must be available in the right place at the right time to help answer increasing complex biological and clinical questions. caBIG® is based on widely accepted industry data standards that enable data "liquidity" among all stakeholders.

Connecting to the Grid—At the end of 2009, there were more than 120 organizations connected to caGrid, hosting more than 125 analytical and data services that were accessible to the community. Increasing quantities of data, including biospecimens, microarray experiments, and other research data are being made available, and more and more researchers are turning to caGrid to find the data they need. Vast quantities of data are also being generated by programs like The Cancer Genome Atlas project and these data are also available via caGrid. As more data is made available, and more organizations connect to the grid, the value of those connections increases dramatically. ■

Mobile Applications

As demands for real-time data access arise, caBIG® and its services-based infrastructure is evolving. In 2009, the NCI and members of the caBIG® community developed three freely available mobile applications to access caBIG® data resources. For more information about caBIG® mobile applications, visit www.cabig.cancer.gov.

caBIG® Links the Cancer Community

The "connected" cancer community continued to grow in 2009, as data interoperability became more widely accepted in support of collaborative cancer research. More and more data sharing occurred as researchers in proteomics, nanotechnology, genomics, imaging, pathology/biorepositories, drug development, and other disciplines sought additional data resources to enhance and accelerate their work.

Today, caBIG® connects the U.S. cancer community, including more than 56 National Cancer Institute (NCI)-designated Cancer Centers, 16 members of the NCI Community Cancer Center Program, NCI SPOREs (Specialized Programs Of Research Excellence), NCI Cooperative Groups, community care facilities, academic institutions, and commercial enterprises. These U.S. organizations are joined by an expanding group of international collaborators in more than a dozen different countries, including the United Kingdom, China, Mexico, Chile, Uruguay, Argentina, Brazil, The Netherlands, Germany, the Czech Republic, Finland, Jordan, India, Pakistan, Australia, and New Zealand.

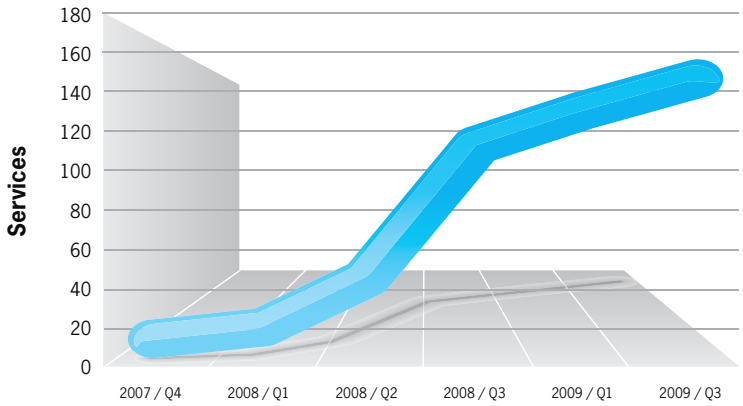
As more organizations connect to caGrid—120 as of December 2009—the value of the data on the network increases.

READ ON OR VISIT US ONLINE AT: CABIG.CANCER.GOV/LINKS

Center Deployment Leads

At the forefront of connecting the cancer community are individuals known as Center Deployment Leads—researchers, bioinformaticians, and IT professionals working at cancer research centers around the U.S. who are actively deploying caBIG® technologies and capabilities. Substantial momentum for the use of caBIG®-enabled data interoperability is directly attributable to the dedicated efforts of Center Deployment Leads within their respective cancer centers.

Connecting the Cancer Community via caGrid



Since the NCI first deployed caGrid in 2007, a growing number of data and analytical services, hosted by the NCI and a diverse collection of organizations, have been made available to researchers across the globe.

There are many different approaches or "pathways" to data interoperability, and each organization can develop its own unique hybrid mix of tools and capabilities, depending on its local needs and stage of digitalization of operations. Some examples include:

- To initiate electronic data-sharing within its own laboratories or departments, an organization can adopt and install existing caBIG® tools to meet its research and data management needs. If there are few legacy systems in place, or limited resources to develop custom solutions, this approach can quickly get an organization on the way to creating and sharing interoperable data.

- To maximize the value of previously developed in-house or 3rd-party tools, an organization can make use of caBIG® Application Programming Interfaces (APIs) and Software Development Kits (SDKs), along with guidance for in-house developers to modify existing products, or develop new applications to be interoperable with caBIG®. This approach works well if an organization has legacy systems and/or significant in-house development resources to create custom software tailored to its specific needs.

- An organization with a large collection of biospecimens may adopt caTissue, providing an integrated environment to manage and track all its biospecimens regardless of the repository location. The same organization may already have a commercial Clinical Data Management System (CDMS) in place to collect and manage data from its clinical trials. In this case, the organization may work with the software vendor, the Clinical Trials Management Software (CTMS) Knowledge Center, or a caBIG® Support Service Provider to adapt the existing CDMS software to interoperate with caBIG®, facilitating data exchange with other applications.

- An organization may implement caGrid technology to connect multiple databases or make data from diverse legacy systems interoperable, providing researchers with a simple way to query for information regardless of where that information is stored. This approach may be applied within a department, across multiple divisions of the organization, or even connecting multiple sites, as Ohio State has done for the Ohio Perinatal Research Project *(see page 16 for more information)*.

- Many organizations make use of more than 75 biomedical terminologies and vocabularies developed and maintained by caBIG® (http://nciterms.nci.nih.gov) to manage and annotate their collected data, making it more available to wide audiences of researchers.

- As the repertoire of NCI Enterprise Services *(reusable software modules with specific functionality, described in the Connectivity section on page 28)* continues to grow, developers can more quickly and easily develop customized services-based applications to handle unique data management needs, while ensuring interoperability with a large community of other organizations making use of caBIG® tools and technology.

Regardless of a particular organization's approach to data interoperability, the caBIG® program provides flexible, modular support resources to assist in the process.

Facilitating Broad Use of caBIG® Tools and Technology

To support the growing caBIG® community, the NCI's Center for Biomedical Informatics and Information Technology (CBIIT) established the **E**nterprise **S**upport **N**etwork **(ESN)**. The ESN is a collection of diverse organizations that provide technical services, mentoring, and expertise to researchers and IT staff. Two components of the ESN include Knowledge Centers and Support Service Providers:

Knowledge Centers

Knowledge Centers were established at institutions with demonstrated expertise in specific areas of interest to caBIG® users. These Knowledge Centers provide support services, including web-based education and domain expertise in tool integration and use, and serve as the interface to the community for ongoing software development requests.

The six Knowledge Centers cover the following areas of expertise:

- **caGrid:** The Ohio State University and The Ohio State Comprehensive Cancer Center, with the University of Chicago and the Argonne National Laboratory
- **Clinical Trials Management Systems:** Duke University Comprehensive Cancer Center, with Northwestern University, Cancer and Leukemia Group B – Information Systems (CALGB-IS), and SemanticBits
- **Molecular Analysis Tools:** Columbia University Herbert Irving Comprehensive Cancer Center with the Broad Institute of MIT and Harvard
- **Tissue/Biospecimen Banking and Technology Tools:** Siteman Cancer Center, Washington University at St. Louis
- **Vocabulary:** Mayo Clinic
- **Data Sharing and Intellectual Capital:** The University of Michigan

Support Service Providers

Support Service Providers are licensed by the NCI and assist clients wishing to install, modify, or better use caBIG® tools and technology, under negotiated client-provider business arrangements. caBIG® Support Service Providers, who range from small software companies to Fortune 500 companies, offer: **Help desk support, Adaptation and enhancement of caBIG®-compatible applications, Deployment support for existing caBIG® applications**, and **Documentation and training materials and services**. For more information about Support Service Providers, please visit: https://cabig.nci.nih.gov/esn/service_providers.

Data Sharing Policies for Diverse Fields of Study

Policies that promote broad-based information sharing under appropriate access terms are needed for compliance with HIPAA (Health Insurance Portability and Accountability Act), government regulations, and institutional policies. caBIG® recognizes that decisions to share data are made locally by organizations that are stewards of the data. To address multiple legal, policy, ethical, cultural, and security issues affecting collaboration, caBIG® participants developed the **D**ata **S**haring and **S**ecurity **F**ramework **(DSSF)** and items such as model agreements, guidance documents, and decision support tools. The purpose of the Framework is to provide the decision logic for analyzing data sensitivity to determine the appropriate data sharing mechanisms and the appropriate security and data access controls to apply to access the data. For more information, please visit: https://cabig-kc.nci.nih.gov/DSIC/KC/.

The Framework and accompanying tools recognize there are varying levels of sensitivity of health information and that many data exchanges require agreements, user validation, authorization of intended uses, etc. Institutions determine who is authorized and under what conditions. The intended audience for the Framework includes researchers and their study teams, IRB members, privacy officers, grant staff, technology transfer staff, and legal counsel. ■

caBIG® in ACTION

The power of caBIG® derives from its ability to bridge people and knowledge from different fields of cancer research and care. Genomics, proteomics, radiology, clinical research, and other areas of study inform one another, and ultimately inform clinical decision-making. In doing so, research and health care improve more quickly, building upon cross-functional knowledge previously trapped in silos.

READ ON OR VISIT US ONLINE AT: CABIG.CANCER.GOV/INACTION

In 2009, the caBIG® community supported a wide variety of research efforts, including new models of clinical and basic research, population science, biospecimen management, and radiology, in many different types of cancer.

New Models of Clinical Research

More than 20 organizations, including **Baylor College of Medicine**, **Georgetown University Lombardi Cancer Center** and **Virginia Commonwealth University**, are using individual tools or combinations of tools from the caBIG® Clinical Trials Suite to address their respective clinical trial needs. A few specific examples are given in the following paragraphs.

Clinical researchers and trial coordinators at the **Kimmel Cancer Center** of **Thomas Jefferson University** improved their clinical trials processes, from protocol development through regulatory submissions. Using caBIG® Clinical Trials Suite, the institution reinvented its participant registration and scheduling, data storage, translation and exchange, and adverse event reporting. As a result, clinical trial costs were reduced and data accuracy was increased.

The Kimmel Cancer Center used the Patient Study Calendar (PSC) from the caBIG® Clinical Trials software suite in four separate trials to manage calendars and track care for 600 participants. In contrast with previous systems, trial coordinators now monitor and manage calendars remotely, and more detailed reports are available to Principal Investigators. Future versions of the Patient Study Calendar will allow connection with Microsoft Outlook™ or Google Calendar™ so appointments and compliance requirements can be easily sent to patients.

The **Mayo Clinic Cancer Center** in Rochester, Minnesota, also used PSC software in a pilot program for multi-site clinical trials. These capabilities now help the center coordinate calendars across hundreds of clinical trial sites, dramatically reducing time and costs of trial management compared to previous methods, resulting in clearer protocols and improved treatment compliance by patients.

PSC was originally developed by **Northwestern University**. In 2009, the Northwestern Memorial Physicians Group (the largest primary care group in Chicago) expanded the capabilities of PSC. The Group modified PSC to build a support application that provides checklist and workflow guidance to assist patients and their physicians with complex schedules of individual cancer treatments, making appointment scheduling more efficient and improving treatment compliance.

Washington University Connects Systemwide: caBIG® Provides Pathway to Enterprise Interoperability

Washington University in St. Louis (WUSTL) has made a commitment to interconnect medical records, laboratory results, and patient outcomes information across 13 member hospitals in the BJC Healthcare system through the creation of the Clinical Investigation Data Exploration Repository (CIDER). Data interoperability is a critical enabler of this project and more than 80,000 data terms across the disparate sources have been harmonized, making CIDER data interoperable with caBIG®. WUSTL has been deeply involved with the caBIG® program since its early days, spearheading development of caTissue, and deploying caTissue and caArray to support researchers across the university. More than 460,000 biospecimen samples in caTissue and 180 microarray experiments in caArray are currently available via caGrid. The staff at WUSTL plans to connect the CIDER data warehouse to their internal version of caGrid and to provide secure, integrated access to data stored in caTissue and caArray for researchers across the campus. Mark Watson, M.D., Ph.D., states, "One key thing that will really accelerate research and our ability to leverage all of these biospecimens is to put them into an information network, where scientists as a group can know what's available for research purposes."

caBIG® Developing a Unified Clinical Data Management Approach

The leadership at the **Winthrop P. Rockefeller Cancer Institute** at the **University of Arkansas for Medical Sciences (UAMS)** is implementing its vision to connect clinical researchers across the entire cancer center using a caBIG® integrated framework. They began by developing a single sign-on dashboard linking multiple tools from the caBIG® Clinical Trials Suite to provide researchers with simple one-stop access to clinical trials data generated at their center. The dashboard integrates data on patient registration (C3PR), treatment calendars (PSC), adverse events (caAERS), and clinical laboratory results (LabViewer). Over the course of 2010, access to biospecimen data from caTissue and microarray gene expression data from caArray, both already in use at UAMS, will also be made available through this integrated interface. Their long-term plans are to drive broader adoption of caBIG® technology beyond the cancer center to include other biomedical researchers at UAMS. States Kari Cassel, Chief Information Officer at UAMS, "caBIG® has allowed us to complete our vision for clinical research and provided a level of robustness in some of the toolkits that have come from the caBIG® collaborative that would've taken us years to develop on our own."

Clinical Trial Adverse Event Reporting

Using the caBIG® Adverse Event Reporting System (caAERS) in a production pilot on nine clinical protocols, the **Mayo Clinic** saves time by leveraging patient information from existing treatment records, automating a process formerly done manually.

caAERS also offers direct reporting to ADEERS, the National Cancer Institute's Adverse Event Expedited Reporting System, improving reporting accuracy by reducing manual data entry.

10 Cancer Centers, including **City of Hope Medical Center** and **Johns Hopkins University** have deployed or are installing caAERS to address their adverse event reporting needs, and another 10 organizations are currently evaluating its use.

Clinical Participant Registry

In 2009, **Duke University Comprehensive Cancer Center**, **Georgetown University**, and **Wake Forest University** participated in PANVAC, a Phase II vaccine study for colorectal patients with liver metastases. The three organizations used caBIG® Central Clinical Participant Registry (C3PR) to coordinate participant enrollment across their sites.

C3PR software allows researchers to register trial participants electronically at trial inception, securely capturing and tracking a wide range of information, including consent, eligibility criteria, stratification, randomization, and screening—all critical capabilities for multi-site studies.

The **Duke-Wake Forest-Georgetown** PANVAC trial coordinates C3PR registration with caBIG® Cancer Central Clinical Database (C3D), which helps ensure FDA-compliance of clinical trials. Registration and ongoing data collection, management, and reporting to the FDA take place electronically from each of the three trial sites.

caBIG®: Enabling Adaptive Clinical Trials at UCSF

The **University of California**, **San Francisco (UCSF)** is collaborating with the National Cancer Institute, the Food and Drug Administration, and several pharmaceutical and biotechnology companies to improve breast cancer treatment by identifying new, faster approaches for assessing the effectiveness of therapeutic treatments. The **I-SPY TRIALs** (1 and 2), more formally known as the **I**nvestigation of **S**erial studies to **P**redict **Y**our **T**herapeutic **R**esponse with **I**maging and molecular **A**na**L**ysis, use a combination of pharmaceutical and technological approaches to enhance clinical trials.

I-SPY 1 was a national study to identify genetic biomarkers that predict response to therapy for women with late-stage breast cancer. Initial findings from I-SPY 1, published in 2009[1], showed significant progress in establishing new Magnetic Resonance (MR) imaging standards and validating the use of new tools for managing biospecimens and clinical trial automation.

I-SPY 2 advances those results by stratifying patients into different treatment options according to their genetic profiles and predicted response to treatment. The new trial is designed to more efficiently screen promising agents and reduce the 15 to 20 year drug assessment process to just a few years. The study is an "adaptive design" in which clinical data is used in real time to direct the course of treatment for trial participants. (http://ispy2.org/)

A sophisticated bioinformatics infrastructure was required to enable the rapid collection, management, and analysis of genomic and outcomes data from I-SPY trial participants. **TRANSCEND** (**TRAN**slational Informatics **S**ystem to **C**oordinate **E**merging Biomarkers, **N**ovel Agents, and Clinical **D**ata) is a scalable framework developed by UCSF that uses caBIG® tools such as caBIG® Integration Hub, caTissue, caArray, and caIntegrator to collect data for I-SPY 2. In addition to automated data collection, the TRANSCEND suite supports patient randomization through an automated web service, as well as automated population of clinical report forms from Electronic Health Records (Tolven).

An Integrated National Clinical Trials Repository

The **C**linical **T**rials **R**eporting **P**rogram **(CTRP)**, which came online in 2009, aims to provide a single, comprehensive, easily accessible database with regularly updated information on all NCI-funded interventional clinical trials. Led by the NCI's Center for Biomedical Informatics and Information Technology, in partnership with the Cancer Therapy Evaluation Program (CTEP), the Division of Cancer Prevention (DCP), and the Center for Cancer Research (CCR), this resource supersedes the many clinical trials databases maintained in disparate parts of the NCI. The electronic infrastructure enabled by caBIG® technology allows grantees to enter information about the status of their clinical studies. The resulting real-time connectivity permits extramural investigators to participate more directly in NCI's clinical trials prioritization process, fulfilling a key recommendation of the NCI-commissioned Clinical Trials Working Group.

[1] *Olson, JA, et. al. 2009. Improved surgical outcomes for breast cancer patients receiving neoadjuvant aromatase inhibitor therapy: results from a multicenter phase II trial. J Am Coll Surg 208(5):906-14. http://www.ncbi.nlm.nih.gov/pubmed/19476859*

Managing Biospecimens

The Siteman Cancer Center at **Washington University** in St. Louis, Missouri, has been a leading developer of the caTissue Suite, and deployed caTissue to manage four major biobanks across the institution. This capability allows investigators to directly query remote biospecimen banks. As of December 2009, data on more than 450,000 biospecimens was available via caGrid.

Thomas Jefferson University customized caTissue, providing integrated search and sample ordering across multiple repositories within the university. caTissue helped the university standardize its biospecimen management by automating the collection of clinical and pathology data, diagnoses, and sample annotation, simplifying research inquiries and eliminating many manually intensive steps.

Vanderbilt University is implementing caTissue to share almost 900 breast and colon cancer samples in 2010. The university will be migrating 20 years of biomedical sample collection into caTissue in order to make the resource broadly available to the cancer community.

Similarly, the **H. Lee Moffitt Cancer Center**, Tampa, Florida, is collaborating with **Ponce University** in Puerto Rico to create an island-wide repository for biospecimens.

Biospecimen management is also important to the 16 NCI **SPOREs** (**S**pecialized **P**rograms **O**f **R**esearch **E**xcellence), which encourage research into particular cancer types. caBIG® technology is enabling research projects across multiple SPOREs. For example, caGrid and caTissue are being used to share information about biospecimens among 11 Prostate SPORE member organizations.

Supporting Basic Research

The **Jackson Laboratory Cancer Center**, Bar Harbor, Maine, uses mouse genetics and genomics to develop disease models for cancer in humans. Jackson Labs has been using caArray to support its own researchers, as well as for sharing data with the broader cancer community. At the end of 2009, the lab's data service on caGrid made arrays from 16 experiments available, containing 369 hybridizations. Jackson Labs is also developing a caGrid-enabled Mouse Phenome Database as a central repository for mouse phenotype data, has harmonized ontologies describing mouse and human anatomy, and will be relying on caGrid for seamless integration of mouse preclinical data with human data.

29 organizations are using caArray to manage their microarray gene expression data, and two thirds of those are sharing that data publicly using caGrid. An additional 41 centers are in the process of installing caArray for internal use. The NCI also maintains a publicly accessible installation of caArray and more than 400 individuals, representing over 220 organizations, have access to a collection of 25,000 microarray experiments.

> "I view caBIG® as being absolutely essential to our strategic mission. We have ambitious plans for the implementation of an integrated clinical and molecular database that can ultimately guide personalized medicine."
>
> Louis M. Weiner, M.D.
> Director
> Georgetown Lombardi Comprehensive Cancer Center

In addition to caArray, the caBIG® program has developed or adapted a diverse collection of software tools to support basic research. These programs include **geWorkbench** (an integrated platform for gene expression, sequence, and pathway analysis), **GenePattern** (gene expression and other genomic data analysis), **caB2B** (for querying multiple data types across caGrid), and **caNanoLab** (developed to support the NCI Nanotechnology Alliance). These tools are deployed across the NCI cancer community and are used at many institutions, including **Baylor College of Medicine**, **Fox Chase Cancer Center**, **The Burnham Institute**, **The Salk Institute**, and **The University of North Carolina Lineberger Cancer Center**.

Managing Biomedical Images

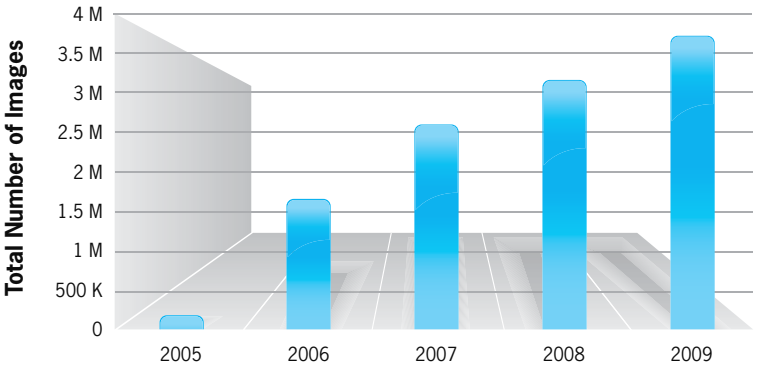
Diagnostic medical images are essential tools for cancer research and care. Radiology and pathology images, now generated digitally rather than on film, provide both challenges for data management and opportunities for sharing, analysis, and integration with other research and clinical data.

The NCI created the National Biomedical Imaging Archive (NBIA) as a searchable online repository for *in vivo* DICOM-format medical images. The NCI maintains a local version of NBIA that researchers may use to store and manage their medical images. Organizations that generate large numbers of images are also choosing to install their own copy of NBIA. Numerous organizations are in pilot or production use of NBIA at their own facilities, including **Washington University at St. Louis**, the **National Cancer Research Institute (U.K.)**, **Thomas Jefferson University**, **Vanderbilt University**, **Christiana Care**, and **Novartis**.

In addition to NBIA, the NCI has developed several unique image-analysis tools to help researchers integrate medical images into their research workflows. Annotation and Image Markup (AIM) software allows radiologists to include standards-based annotations along with the images, simplifying data exchange

between different organizations. The eXtensible Imaging Platform (XIP) is used by researchers and commercial caBIG® partners such as Siemens Research. XIP provides a toolkit of standards-based components to construct image analysis applications tailored to researchers' specific needs, facilitating multi-site data collection and comparison.

Images Hosted at the NCI



More than 1,100 individuals representing over 800 organizations have registered for access to the National Biomedical Imaging Archive (NBIA) hosted at the NCI, which currently contains more than 3.7 million medical images. That number is steadily increasing over time.

Using caGrid to Connect the Dots

Numerous Cancer Centers are using caGrid technology to connect researchers within their own organizations, as well as connecting with external organizations. The flexibility of caGrid is leading to innovative solutions to address diverse data management tasks.

MyGrid and Taverna

caBIG® researchers at the U.S.-based Argonne National Laboratory have been collaborating with U.K.-based researchers from the MyGrid project. Together, the two organizations worked to improve research workflows, leading to the adoption of TAVERNA technology (http://www.taverna.org.uk/) to create modular, reusable workflows for a collection of caBIG® tools (http://cabig.cancer.gov/resources/newsletter/issueXXII/action.asp). By developing reusable workflows, researchers can improve the quality and reproducibility of their data.

caBIG® Across the Enterprise in Alabama

The University of Alabama at Birmingham (UAB) won the 2009 caBIG® Deployment Award for its innovative approach to developing a common IT architecture. The team at UAB is using caBIG® technology to drive an enterprise-wide approach to data integration. They began by performing a detailed analysis of their existing and planned IT needs and determined that caGrid could link a wide variety of systems from Electronic Health Records, to billing, to clinical systems and tissue banks across multiple communities within the university. In this way, caBIG® technology is making it much simpler for researchers and clinicians both on-campus and off-campus to securely access data from multiple resources, and find the information they need regardless of where the data are stored.

UAB researchers currently have access to caTissue, caArray, and the Computational Portal and Analysis System (CPAS), for managing research data, with wider implementation of caBIG® tools and technology planned to address additional data management needs. As one of their first integration projects, the UAB team enabled access to clinical outcomes data housed at the Medical Center for translational researchers in the university's basic research labs. John Sandefur, Information Systems Manager at UAB says, "In the short term, the enormous volume of clinical data is the "bait" to catalyze collaboration between scientists. The grid will drive more collaboration as researchers see the benefits of data sharing to their own work, and as we get more data on the grid, the value continues to increase."

caBIG® Here, There and Everywhere Across OSU

The **Ohio State University (OSU)** has a long association with the caBIG® program. Researchers at OSU have been instrumental in developing caGrid, and OSU currently hosts the caBIG® caGrid Knowledge Center. It comes as no surprise, therefore, that OSU has looked to caGrid technology to connect a wide variety of diverse systems and departments across the entire university, and beyond. Phillip Payne, Ph.D., Director of the Biomedical Informatics Program at OSU states, "We really want to find a way to link this very robust middleware developed by the NCI with a lot of interesting clinical and translational research use cases, at the community level, at the national level, and locally within our institution."

As an initial step toward broader deployment within OSU, the OSU team has leveraged a Clinical and Translational Science Awards (CTSA) grant from the NIH to develop web portals that enable data integration and facilitate collaborative translational research within its own walls. On a broader scale, the university is applying caGrid to handle data access and federation challenges at the Cancer Center, and across all 17 colleges of OSU.

Externally, OSU is collaborating with the University of California, San Francisco and several other CTSA-funded institutions on the Human Subjects Study Database Project. The collaboration is creating a federated queryable repository for metadata that describes the clinical trials being conducted by all of the CTSA-funded sites. By applying caBIG® technology to manage the data federation, each institution can retain control of its own data and selectively determine which data is accessible to other organizations.

The OSU team is also applying caBIG® infrastructure components to connect hospitals in the OSU system with the Nationwide Children's Hospital in Columbus, Ohio. As part of the Ohio Perinatal Research Program, the ambitious effort was launched to better understand and address medical issues around premature birth and high infant mortality in traditionally underserved populations in Columbus. Rather than creating a single massive data warehouse, OSU staff are using caGrid to link the already well-developed data repositories at both institutions. OSU staff have leveraged caBIG®-controlled vocabularies to map common data elements between the data collections. In 2010, integrated web portals will allow researchers at either institution to query more than 600 phenotypic variables describing key clinical and biospecimen-related data on both mothers and children.

> "One of the biggest issues we have is with databanks. Getting those databanks to talk to each other is the real key. caBIG® has allowed us to work more globally and allowed us to communicate within those systems. The technology of caBIG® is going to be a tremendous advantage to moving cancer care forward."

Nicholas Petrelli, M.D.
Bank of America Endowed Medical Director
Helen F. Graham Cancer Center, Christiana Care Health Systems

Connecting Research and Care in the Community

The **NCI Community Cancer Centers Program (NCCCP)** uses a national network of community cancer centers to expand cancer research and deliver the latest, most advanced cancer care to Americans in their home communities. To help achieve these goals, the 16 participating NCCCP sites are sharing specimens, research data, and clinical outcomes data using caBIG® technology.

At the forefront of caBIG® adoption is the **Helen F. Graham Cancer Center at Christiana Care Health System** in Delaware. In addition to participating in the NCCCP program, Christiana Care is also a founding member of the Center for Translational Cancer Research (CTCR) in Delaware, formed in collaboration with the University of Delaware, the Nemours/AI duPont Hospital for Children, and the Delaware Biotechnology Institute. To support its diverse translational research efforts, Christiana Care has implemented caTissue to improve management of its biospecimens. In 2010, the Cancer Center is planning to integrate biospecimen data with gene expression data housed in caArray and medical image data housed in NBIA.

Other community cancer centers deploying caBIG® tools include **Our Lady of the Lake Regional Medical Center** in Baton Rouge, Louisiana. The medical center has deployed NBIA for managing medical images within its own facility and at its sister-site, **Mary Bird Perkins Cancer Center**, also in Baton Rouge. In addition, two centers from the **St. Joseph's Healthcare Network** are in the process of installing caTissue to support their clinical care efforts. A number of additional NCCCP members are planning to install caBIG® applications over the course of 2010 to address their own specific research challenges.

Additionally, NCCCP sites are participating in the caBIG®-ASCO collaboration to design an oncology-extended Electronic Health Record that will utilize caBIG® standards for interoperability.

The Next Generation of Research: *In Silico* Centers

In 2009, a series of caBIG® "*In Silico* Research Centers of Excellence" were designated by the NCI. The contracts were created to support investigator-initiated, hypothesis-driven research into cancer using bioinformatics and data mining techniques with a focus on mining collections of publicly available data such as The Cancer Genome Atlas. The *In Silico* Research Centers are organized into a national consortium that promotes collaboration amongst the Centers, and between the Centers and the cancer research community. Contracts were awarded to **Emory University**, **Georgetown University**, **Fred Hutchinson Cancer Research Center**, **Columbia University**, and the **Translational Genomics Research Institute (TGEN)**.

In addition to generating and publishing novel scientific findings obtained through the use of caBIG® and other bioinformatics tools, it is expected that new workflows and new tools to utilize these publicly available data stores will be developed by the *In Silico* Centers. These tools and all experimental results will be shared with the broader research community in a format interoperable with caBIG®.

The first novel work from an *In Silico* Research Center was published in *Nature* in January 2010 by **Columbia University**[2]. In 2009, Dr. Andrea Califano's group studied glioblastoma data from the TCGA project and examined gene expression profiles, comparative genomic hybridization, and copy-number data to identify two genes, C/ERPßβ and STAT3C. The genes were specifically associated with activating genes in the mesenchymal pathway, markers for particularly aggressive GBM tumors. The ability to rapidly perform integrative research *in silico* across large, diverse data sets and make novel associations of specific genes with certain tumor phenotypes clearly demonstrated the value of interoperable tools and data sets. ■

[2]*Carro, MS, et. al. 2010. The Transcriptional Network for Mesenchymal Transformation of Brain Tumors. Nature 463(21):318-327. http://www.ncbi.nlm.nih.gov/pubmed/20032975*

New Models of Collaborative Cancer Research and Care

Researchers are seeking reliable "biomarkers" such as gene expression patterns to help identify cancer subgroups. Such information helps speed the development of targeted therapies to enable physicians to individualize treatment.

The Cancer Genome Atlas Project

The **C**ancer **G**enome **A**tlas **(TCGA)**, a collaborative research project between the **NCI** and the **N**ational **H**uman **G**enome **R**esearch **I**nstitute **(NHGRI)**, was created to apply genome analysis techniques to better understand the underlying causes of cancer. Initially focusing on lung, brain (glioblastoma), and ovarian cancers, the project was expanded in 2009 to include more than 20 cancer types.

READ ON OR VISIT US ONLINE AT: CABIG.CANCER.GOV/NEWMODELS

Member organizations across the country are applying a variety of molecular analysis techniques, including Genome-Wide Association Studies (GWAS) and high-throughput DNA sequencing to drive cancer research. Researchers seek to identify both large-scale genomic changes, including copy number variations and chromosomal translocations, and individual sequence mutations that contribute to cancer development or progression.

Data generated from these collaborative studies is made available through public databases and portals supported by caBIG® technology, including the **TCGA data portal** (http://cancergenome.nih.gov/dataportal) and the **C**ancer **M**olecular **A**nalysis **(CMA)** portal (https://cma.nci.nih.gov/cma-tcga/).

Multidisciplinary Collaboration Portal

A diverse collection of genomic and clinical data from TCGA, TARGET, and other NCI projects is currently available through the **C**ancer **M**olecular **A**nalysis **(CMA)** web portal. Using caIntegrator, the CMA web portal helps researchers collect and correlate clinical outcomes data with genomic data from these data sets. Through the caBIG® program caIntegrator2, the integrative research capabilities of the CMA portal are now locally available to researchers, allowing them to easily create customized translational research web portals for their own study data quickly, without the need for programming skills.

Collaborative Childhood Cancer Treatment Research

The National Cancer Institute's **T**herapeutically **A**pplicable **R**esearch to **G**enerate **E**ffective **T**reatments **(TARGET)** program employs genomics technologies to rapidly identify therapeutic targets in childhood cancers. Ten children's hospitals and pediatric research centers linked through caBIG® are TARGET collaborators.

Recently, a group led by researchers at **St. Jude Children's Research Hospital** identified mutations in a gene that predicts a high likelihood of relapse in children with acute lymphoblastic leukemia (ALL)[3]. The gene, called IKZF1, may form the basis for future diagnostic tests if the initial results are validated. An accurate molecular test to identify this prognostic biomarker may help physicians prescribe the most appropriate therapy for suitable patients.

Collaborative Brain Cancer Research

The **RE**pository of **M**olecular **BRA**in **N**eoplasia **D**a**T**a **(REMBRANDT)** project, run by the **Glioma Molecular Diagnosis Initiative**, is focused on uncovering the underlying causes of *glioblastoma multiforme* (GBM) cancer. REMBRANDT provides a bioinformatics framework and web portal that allows researchers to integrate clinical and functional genomics data from clinical trials involving patients suffering from Gliomas[4]. For example, researchers can query across these data sets and identify relationships between gene expression and clinical outcomes. The program has collected data from more than 550 patients to date.

A second project, **VASARI** (**V**isually **A**cce**S**s**A**ble **R**embrandt **I**mages), is an extension of the original REMBRANDT trial. The VASARI study seeks to validate the use of medical images as predictive biomarkers for cancer diagnosis. MRI images on the same samples used in the REMBRANDT program are used in the VASARI trial. Images are managed through the National Biomedical Imaging Archive (NBIA).

[3] *Mullighan, CG et.al. 2009. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med. 360(5):470-80 http://www.ncbi.nlm.nih.gov/pubmed/19129520*

[4] *Madhavan, S et. al. 2009. Rembrandt: helping personalized medicine become a reality through integrative translational research, Mol Cancer Res. 7(2):157-67 http://www.ncbi.nlm.nih.gov/pubmed/19208739*

A third project, NCI's **C**ancer **GE**netic **M**arkers for **S**usceptibility **(CGEMS)**, has conducted genome-wide association studies (GWAS) to identify common gene variations that influence risks for cancer. Over 550,000 common genetic variants were analyzed by the end of 2009 and the data was made available through caGWAS (caBIG® Genome-Wide Association Studies) software. The results have been aggregated and made available using caIntegrator.

Collaborative Epidemiology Research

The NCI Division of Cancer Control and Population Sciences (DCCPS) supported several projects in 2009 using caGrid and controlled vocabularies. These projects demonstrate the adaptability of caBIG® technology for population science research and provide an essential component of the broader DCCPS vision for integrating a broad collection of public health data through the use of interoperable informatics over the next few years.

One of the initiatives was the Population Science Grid (PopSciGrid) 1.0 project, a pilot project developed at **Northwestern University** in collaboration with the NCI, and completed in early 2009. Using geographical data about smoking prevalence and cigarette taxes, researchers were able to query diverse data types housed at geographically dispersed sites and connected via caGrid. Version 2.0 is currently under development and dramatically expands the capabilities of the original project, leveraging a broad collection of data sources, including public surveillance, grants data, and remote sensing information and developing a set of data analysis tools to enable remote access to the information.

Since 1976, the National Cancer Institute has gathered information on cancer incidence and survival, covering approximately 26 percent of the U.S. population with cancer in its SEER (Surveillance, Epidemiology and End Results) database. The caSEER project is developing ways to provide a caBIG®-compatible interface with web-based and programmatic access to the SEER data.

The Grid-Enabled Measures (GEM) database, also developed by researchers at DCCPS, enables sharing of population science measures, including behavioral data, on caGrid. This capability allows researchers to search for constructs and related measures, access metadata about these measures (e.g., validity, reliability, constructs, history, and usage), download them, and also submit alternative measures, all using harmonized common data elements. In this way the community actually using these measures would be able to evaluate, apply, and improve existing measures, and share that information easily with the broader community, leading to improvements in data collection and increasing collaboration. Data curation began in 2009 and the GEM database will be available on caGrid in 2010.

Equally intriguing is the PRO-CTCAE project, which aims to develop an electronically based system for patient self-reporting of serious adverse events (AEs) listed in NCI's Common Toxicity Criteria for Adverse Events (CTCAE). This effort should improve the accuracy and precision of grading of AEs, fulfilling federal requirements and facilitating the evaluation of new therapies. Currently under development, the PRO-CTCAE software system is an open-source, standards-based web application that will be interoperable with applications in the caBIG® Clinical Trials Suite, supporting both the Patient Study Calendar (PSC) and caBIG® Adverse Events Reporting System (caAERS). The PRO-CTCAE will complement the conventional CTCAE and provide a novel entry point for patient-reported data in trials.

Collaborative Proteomics Research

In 2009, a growing number of applications became interoperable with caBIG®. For example, TRANCHE, originally developed by the **University of Michigan**, under the auspices of the National Cancer Institute Clinical Proteomic Technologies Initiative for Cancer, enables researchers around the world to easily locate and access large collections of mass spectroscopy data. TRANCHE supports research in a wide variety of diseases, and as of December 2009, hosted 11 terabytes of data across 16 servers in the United States and Japan, with plans to expand to over 80 terabytes in 2010.

caBIG® Beyond Cancer

caBIG® has always been intended to serve as a model for widespread interoperability and data liquidity across the biomedical enterprise. In 2009, caBIG® technologies were used in a number of research projects focused on diseases other than cancer.

> "caGrid was built by the NCI, but, under the hood, there's nothing about the platform that's closely coupled to only cancer research."

Phillip Payne, Ph.D.
Director, Biomedical Informatics Program
The Ohio State University

Cardiovascular Disease

The **C**ardio**V**ascular **R**esearch **G**rid **(CVRG)** project, under development at **Johns Hopkins University**, **Ohio State University**, and the **University of California at San Diego**, is employing caGrid and adapting caBIG® terminology to enable collaborative research across the cardiovascular research community. Because both caGrid and the CVRG leverage common grid technology, they may be connected directly, effectively becoming a "grid of grids."

Arthritis and Skin Disease

The **O**steo**A**rthritis **I**nitiative **(OAI)** of the **N**ational **I**nstitute of **A**rthritis and **M**usculoskeletal and **S**kin Diseases **(NIAMS)** produces large numbers of knee image scans, which have been shared among researchers by physically mailing computer hard drives. The National Biomedical Imaging Archive (NBIA) database from caBIG® has been adopted by members of the OAI to manage and share these images online, eliminating the need to physically ship hard drives, providing rapid data access and reducing potential data loss.

AIDS

The **N**ational **N**euro-AIDS **T**issue **C**onsortium **(NNTC)** seeks to better understand the action of the AIDS virus in the brain, but it is handicapped by a shortage of high-quality tissue samples. As a result, methods to rapidly share information derived from available samples has become critical. Researchers can now use caArray to manage tissue sample data and share information about molecularly characterized tissue samples across multiple geographic sites. Using caArray in such a collaborative manner was initiated by **Indiana University** and **Purdue University**, following the adoption of caArray at Indiana University's Simon Cancer Center.

> "The value of these tissue samples has increased because now you have more longitudinal data associated with the tissue sample, but the NNTC has fewer samples than it originally expected. The ability to fully utilize the data available is critical to the future advancement of Neuro-AIDS research."

Ganesh Shankar
Manager, Advanced IT Core for the Indiana School of Medicine
Indiana University

(United States, Mexico, Chile, Uruguay, Argentina, Brazil, U.K., Netherlands, Germany, Czech Republic, Finland, Jordan, Pakistan, India, China, Australia, New Zealand)

Organizations across the globe are turning to caBIG® tools and technology to solve their data management challenges.

Collaboration **Across Borders**

Because cancer knows no borders, the caBIG® community actively collaborates with organizations in other countries, exchanging ideas developed here and abroad to advance cancer research and care. Multi-national efforts underway in 2009 included:

United Kingdom

The U.S. **National Cancer Institute** and the U.K. **National Cancer Research Initiative (NCRI)** continue building on a long-standing cooperative effort to support data interoperability. To date, collaboration has focused on use of caGrid technology to connect applications developed by researchers in both countries. Among these efforts is the connection of the NCRI's **ON**cology **I**nformation E**X**change **(ONIX)** portal to caGrid, enabling cancer researchers around the world to access a large number of research databases maintained in the U.S. and the U.K. (http://cabig.cancer.gov/resources/newsletter/issueXXIII/action.asp).

> "We're interested in interoperability and joining the currently isolated islands of informatics resources that already exist. To develop a 'network of networks' is our goal, and with our colleagues at caBIG® we have been successful in starting to do this for real."
>
> Alan Hogg, Ph.D.
> Interim Director
> National Cancer Research Initiative (U.K.)

Latin America

The **Latin American Cancer Pilot Program** began in 2009 with Argentina, Brazil, Chile, Mexico, and Uruguay. In 2010, two clinical trials will be conducted on molecularly characterized stage II and III breast cancer patients in the partner countries. To ensure that cancer researchers can compare data collected at all sites in the study, a broad suite of caBIG® tools will be used to support the project, including caTissue, caArray, and software from the caBIG® Clinical Trials Suite. Patient enrollment is expected to begin in early 2010.

China

The **Duke University Comprehensive Cancer Center** and the **Beijing University Cancer Hospital** launched a pilot collaboration in 2009 using caBIG® tools and technology. Treating more than 150,000 patients per year, the Beijing Cancer Hospital represents a substantial data source for cancer researchers worldwide. This collaboration marks the first clinical trial in China to conduct all patient registrations electronically. The Duke Comprehensive Cancer Center is also helping to facilitate adoption of caBIG® standards in the broader Chinese health system.

India

Introductory meetings were held in 2009 between the National Cancer Institute and Indian ministries responsible for health care. Representatives from the **All Indian Institute of Medical Sciences**, the **Center for Development of Advanced Computing**, and the **Tata Memorial Hospital of Mumbai** participated in these initial meetings. Topics of discussion focused on the use of grid computing for managing clinical trials data.

Additional discussions are underway with institutions in numerous other countries, including Australia, Pakistan, The Czech Republic, Finland, Germany, Jordan, and The Netherlands. ■

Our Role in t
Future of B

A Defining Moment

President Obama has committed to finding a "cure for cancer in our time." The nation's first-ever Chief Technology Officer, Aneesh Chopra, appointed to bring technology-driven innovation to health care and the economy, is seeking ways to "deliver the kind of network capacity….to engage the physician community and the (cancer) care centers, to help us….achieve the same level of analytics we can see in other parts of the economy."

READ ON OR VISIT US ONLINE AT: CABIG.CANCER.GOV/FUTURE

he
iomedicine

It is not surprising that significant changes in the cancer landscape are focused heavily on information technology and informatics. The vast amount of genomic information we now have—increasing daily as a result of high-throughput technologies—threatens to overwhelm even the most sophisticated researcher. And data from the widely anticipated "$1,000 genome" will only add to this tsunami of data.

Other trends have also placed information technology front and center. For example, the U.S. is currently investing more than $44 billion in Electronic Health Records to drive digitization of medicine in order to reduce costs and improve quality of care.

Emergence of the Rapid Learning Healthcare System

In 2007, the Institute of Medicine published a landmark report[5] describing a new approach called the Rapid Learning Healthcare System, in which data of all types is contributed, analyzed, and disseminated in a continuous feedback loop that integrates research and care. Such a system provides the gateway to 21st century medicine—personalized, predictive, preemptive, and participatory—where the patient and his/her particular genetic and clinical characteristics are at the center of all activity, playing a key role in both research and self-management of health.

Cancer as the Model

The prevalence of, and the biomedical infrastructure around cancer, make it an ideal candidate for demonstrating the Rapid Learning Healthcare System, and serving as a model for all of biomedicine. The disease accounts for nearly one out of four deaths in America. There is a standing infrastructure of cancer centers through which treatment is delivered and research is conducted. And cancer research has been at the cutting edge of molecular analyses, leading the way to expedite progress for others.

NCI Strategy

The NCI strategy for demonstrating the Rapid Learning Healthcare System in cancer has three inextricable components: community, connectivity, and content.

Community: In 2008, the NCI established the BIG Health Consortium™, a partnership comprised of all the key stakeholders in health care—consumers, care deliverers, researchers, informatics companies, payers, and government—who agree to work together in a new ecosystem across traditional silos.

"What is it that we can do right now, over the next three to six months, that will make the experience of our cancer researchers and our loved ones who are suffering from cancer, a little bit better, a little bit more impactful? We have it in our capacity, we have the infrastructure, we just have to find a way to come together and make it work."

Aneesh Chopra
U.S. Chief Technology Officer
The White House

[5] *The Learning Healthcare System. Workshop Summary, March 2007 (http://www.iom.edu/Reports/2007/The-Learning-Healthcare-System-Workshop-Summary.aspx)*

The Consortium has identified many of the key attributes of the Rapid Learning Healthcare System, and has planned and launched numerous projects to demonstrate them in real-world settings, including:

- **Athena Breast Health Network**. **The University of California** has launched a statewide collaboration for breast cancer patients to revolutionize the course of their care by designing and testing new approaches to research, technology, and healthcare delivery. The project will initially involve 150,000 women throughout California who will be screened for breast cancer and followed for decades through **University of California** medical centers at **San Francisco**, **Davis**, **Los Angeles**, **San Diego**, and **Irvine**. Among the goals are to create common systems to integrate clinical research and care, and to deliver personalized and biologically targeted care using breast cancer as a prototype. caBIG® is providing guidance on developing an informatics strategy to underpin the Network.

- **Health of Women Study**. The **Dr. Susan Love Research Foundation** and the BIG Health Consortium™ have partnered, in collaboration with the **Beckman Research Institute** at the **City of Hope**, to develop a new model for research called the Health of Women (HOW) Study. The HOW Study is the first-ever online large cohort study of one million women designed to examine causes, treatment, and prevention of breast cancer. Given the cost, complexity, and lengthy time periods associated with the traditional method of designing cohort studies, this approach of engaging consumers early and often—through web-based technology—has the potential to revolutionize how cohort studies are conducted in the future. The HOW Study is leveraging caBIG® infrastructure and web-based software to empower women to connect with the world of research. Replicating the HOW study in other cancers and beyond cancer, using the caBIG® infrastructure "under the hood", will be a rapid and straightforward activity.

"We would not be able to do this project without the collaboration of caBIG®....leveraging a lot of what has been done already....has allowed us to jump start."

Susan Love, M.D.
President
Dr. Susan Love Research Foundation

- **Adolescent and Young Adult Oncology (AYAO) Biorepository Project**. **The Lance Armstrong Foundation** is illustrating how multiple sites can work together in a federated and collaborative way to create an interconnected, virtual tissue bank that focuses on tissue samples from individuals between the ages of 15 to 39 who have been diagnosed with cancer. The Foundation's LIVE**STRONG** (http://livestrong.org/) initiative has funded six sites to inventory their biospecimen samples, and will be leveraging caTissue technology to develop an integrated and universal portal that allows researchers to review, query, and access biospecimen collections across sites.

- **Electronic Patient-Reported Outcomes Project**. Led by the **Duke Comprehensive Cancer Center**, this project will demonstrate personalized patient-centered care and research by using research-quality patient-reported outcomes data to simultaneously inform practice, be incorporated into clinical trials, and tailor patient content. The system will capture patient-reported information through web-enabled interfaces, and will facilitate ongoing patient care, clinical research, and quality evaluation. caBIG® has provided standard data elements to help facilitate data interoperability and liquidity in these clinical and research settings. Such electronic tools to engage consumers and patients can easily be transferred from the cancer setting to all disease management.

Connectivity: To achieve the Rapid Learning Healthcare System, it is essential to make massive amounts of disparate information accessible in real-time based on industry-recognized standards. Such "data liquidity" permits data to be understood in human- and machine-readable forms. caBIG®—having developed the basic infrastructure and tools to conduct clinical research, discovery research, biobanking, and imaging—is now building an electronically connected, virtual biomedical capability called the Cancer Knowledge Cloud, composed of data, applications, and computational capacity. Such a "Knowledge Cloud" can be extended to all disease areas using the same technology infrastructure. Key capabilities include the following:

- caBIG® is implemented through a **s**emantically-aware **S**ervice **O**riented **A**rchitecture **(sSOA)** that supports the integration of diverse data types (e.g., outcomes, clinical encounters, biospecimen annotations, medical images, molecular biology information, etc.) The sSOA enables organizations to facilitate collaborative data processing and work flow execution. These services operate in modular units that can be combined to provide complex data management capabilities, such as enabling meaningful use of Electronic Health Records and quality reporting. The reusable nature of these services speeds the development of novel, interoperable applications through the use of defined, standards-based connections.

- NCI is collaborating with the **A**merican **S**ociety of **C**linical **O**ncologists **(ASCO)** to define the specification for an oncology-extended Electronic Health Record (caEHR) that will facilitate improved care of cancer patients as well as the collection and use of clinical outcomes to drive new knowledge. To date the collaboration has produced the **CORE** (**C**linical **O**ncology **R**equirements for the **E**HR) specification, describing the functions that oncologists want an EHR to perform; providing the structured data elements to be used in oncology EHRs; and defining a common set of interoperability standards that will allow oncology-specific data to be shared from one EHR to another.

Content: Each individual cancer patient represents a "world of information" derived from his/her clinical experience. By combining information from many cancer patients, these data can be converted into knowledge. As part of the Cancer Knowledge Cloud, NCI is launching a Patient Outcomes Data Service based on the terminology and standards of caBIG® and the capabilities of the sSOA. This data service will enable physicians to collect and share information on the cancer diagnostic, treatment, and clinical outcome of individual patients, as well as the outcomes of all their patients in the aggregate. Moreover, such a Data Service can be adopted by other disease communities seeking to gain knowledge by aggregating patient outcomes on a large scale.

The Patient Outcomes Data Service will allow oncologists to do the following electronically:

- Transfer information from the Electronic Health Record about a patient's diagnosis, treatment interventions, and clinical outcome.
- Transfer such a record to the patient within 48 hours of a clinical encounter.
- Compare contemplated treatments with those that are working most effectively for patients within their own clinical practice, across their entire institution, and eventually across multiple institutions nationwide. ■

The caBIG® initiative invites all members of the biomedical community—in cancer and beyond—into its community to participate, share technology, and transform data into knowledge. Specific information can be found at the websites below.

General information about caBIG®:
http://cabig.cancer.gov

Information for basic researchers:
https://cabig-kc.nci.nih.gov/Molecular/KC/

Information for clinical researchers:
https://cabig-kc.nci.nih.gov/CTMS/KC/

Information for biospecimen managers:
https://cabig-kc.nci.nih.gov/Biospecimen/KC/

Information about legal and regulatory topics:
https://cabig-kc.nci.nih.gov/DSIC/KC/

Information about getting connected to caGrid:
https://cabig-kc.nci.nih.gov/CaGrid/KC/

Information about caBIG® technology and software for developers and IT staff: **https://cabig.nci.nih.gov/**

Questions about specific software applications:
http://ncicb.nci.nih.gov/support

Information on participating in the BIG Health Consortium™:
http://bighealthconsortium.org

NIH Publication No. 10-7603
Printed June 2010