National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# Overview of The Cancer Biomedical Informatics Grid™ (caBIG™):

*Connecting the Cancer Community*

NATIONAL INSTITUTES OF HEALTH

NATIONAL CANCER INSTITUTE

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# caBIG™ Introductory Seminars: March 2006

- Topics
  - caBIG™ Overview – March 13
  - Overview of caBIG™ Activities for Clinical Trials and Tissue Banking – March 15
  - Overview of caBIG™ Activities for Integrated Cancer Research – March 16
  - caBIG™ Interoperability and Compatibility Basics – March 17
- https://cabig.nci.nih.gov/seminars
- http://videocast.nih.gov/

National Institutes of Health | National Cancer Institute

# Talk Agenda

- **Introduction and Background**
  - **Program Motivations**
  - **Program Structure and Process**
  - **Program Timeline and Products**
  - **Finding Your Way Around caBIG™**

National Cancer Institute


# Who are we?

- The Center for Bioinformatics is the NCI's strategic and tactical arm for research information management
- We collaborate with both intramural and extramural groups
- Mission to integrate and harmonize disparate research data
- Production, service-oriented organization. Evaluated based upon customer and partner satisfaction.

# Talk Agenda

- Introduction and Background
- **Program Motivations**
- Program Structure and Process
- Program Timeline and Products
- Finding Your Way Around caBIG™

National Cancer Institute

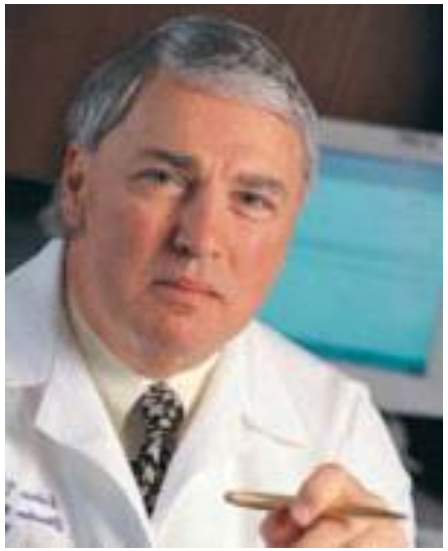
caBIG™ cancer Biomedical Informatics Grid™

# Why Now?

**NCI 2015 challenge goal: eliminate suffering and death due to cancer**

"When I look into the eyes of a patient losing the battle with cancer, I say to myself, It doesn't have to be this way."
*The Nation's Investment in Cancer Research* (2003)

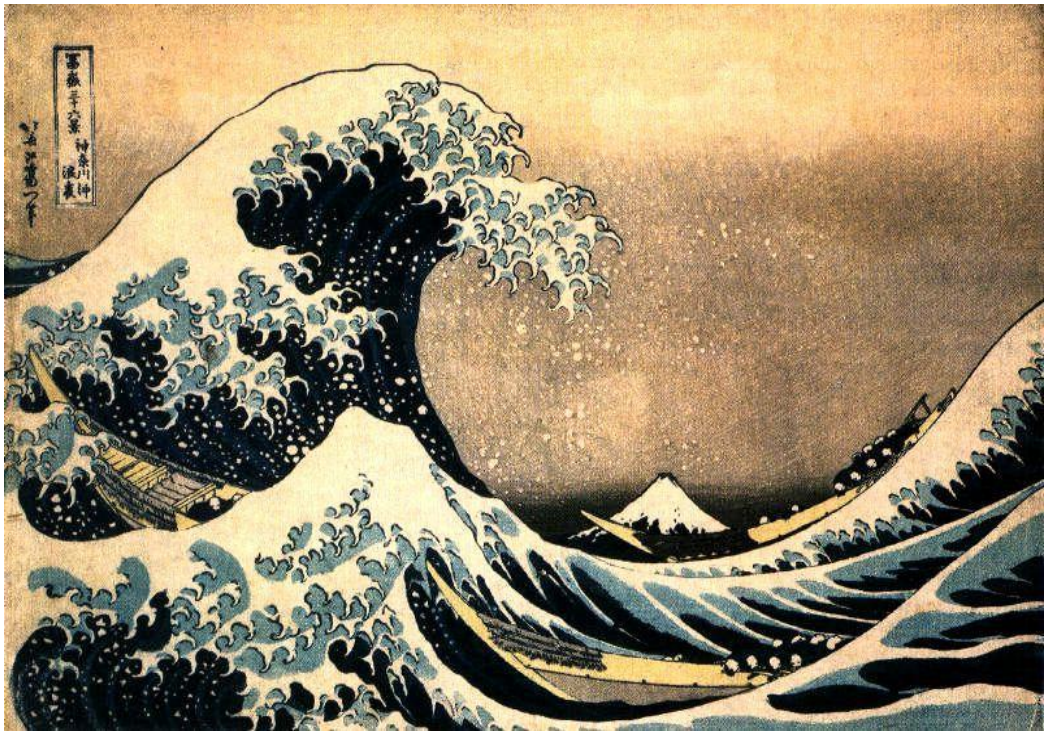
"Nearly every facet of NCI's strategic plan to eliminate suffering and death due to cancer is predicated on the revolutionizing potential of caBIG™."
*Cancer Bulletin, April 12, 2005*



Dr. A.C. von Eschenbach, M.D.
Director, National Cancer Institute

NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

# Health information tsunami



- overwhelming volume of data
- multitude of sources

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# Informatics tower of Babel

- Each part of the health community speaks its own scientific “dialect” (e.g. lab values, genetic profile, clinical data)
- Lack of consensus on common standards and terms
- Lack of coordination across, and collaboration within, the cancer research enterprise
- Integration is critical to achieve promise of molecular medicine

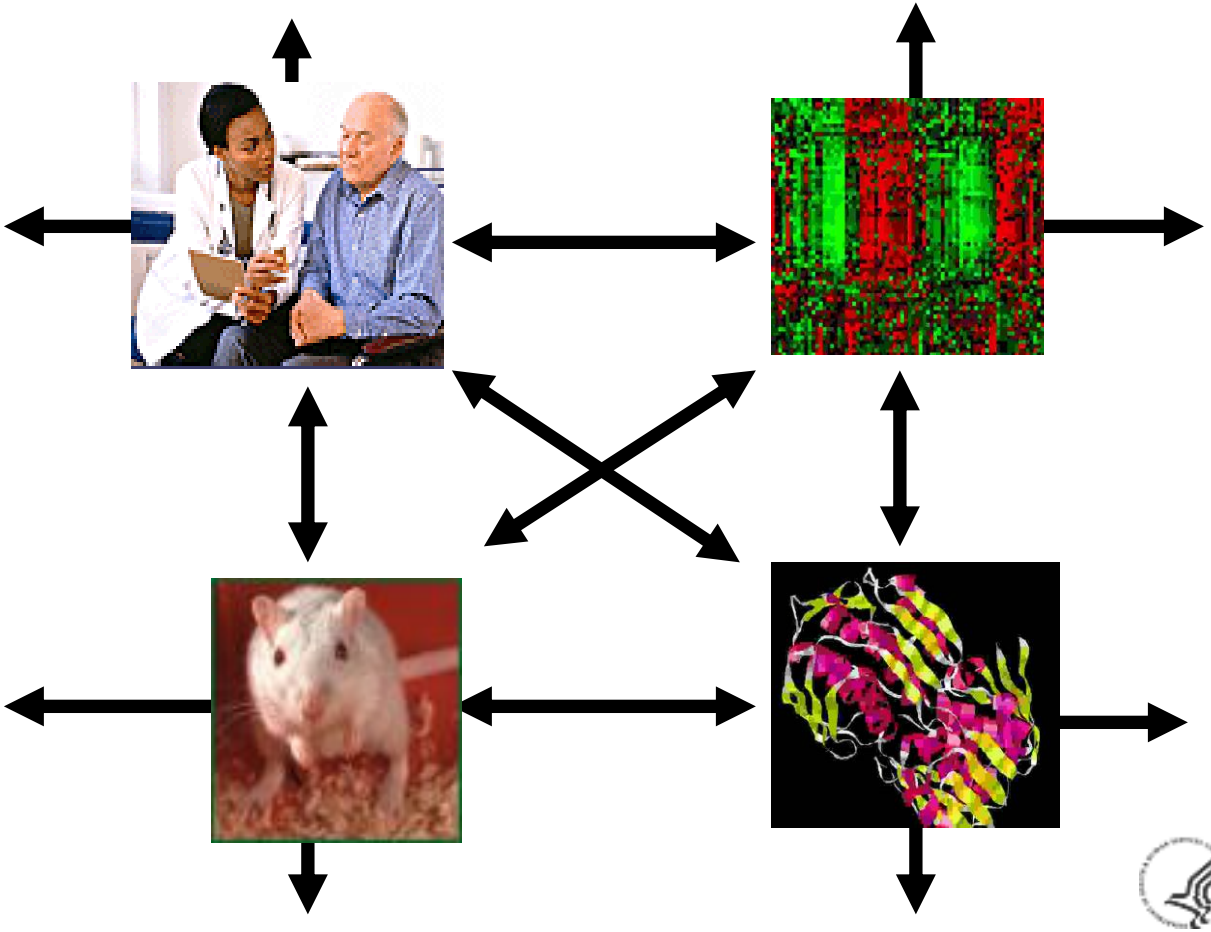
NATIONAL INSTITUTES OF HEALTH

NATIONAL CANCER INSTITUTE

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# caBIG™ infrastructure joins diverse data within an organization and across the country

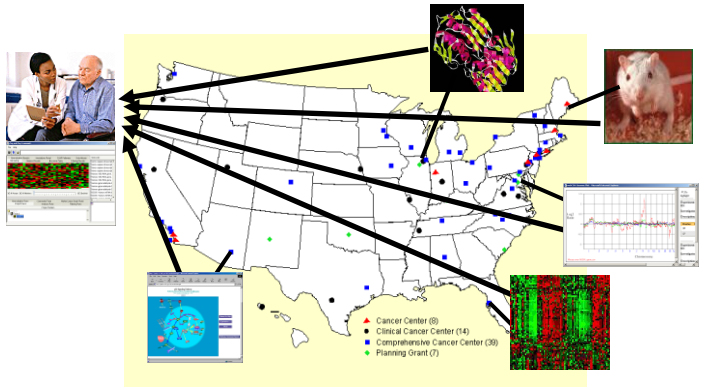


National Institutes of Health

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# Cancer Biomedical Informatics Grid™ (caBIG™)

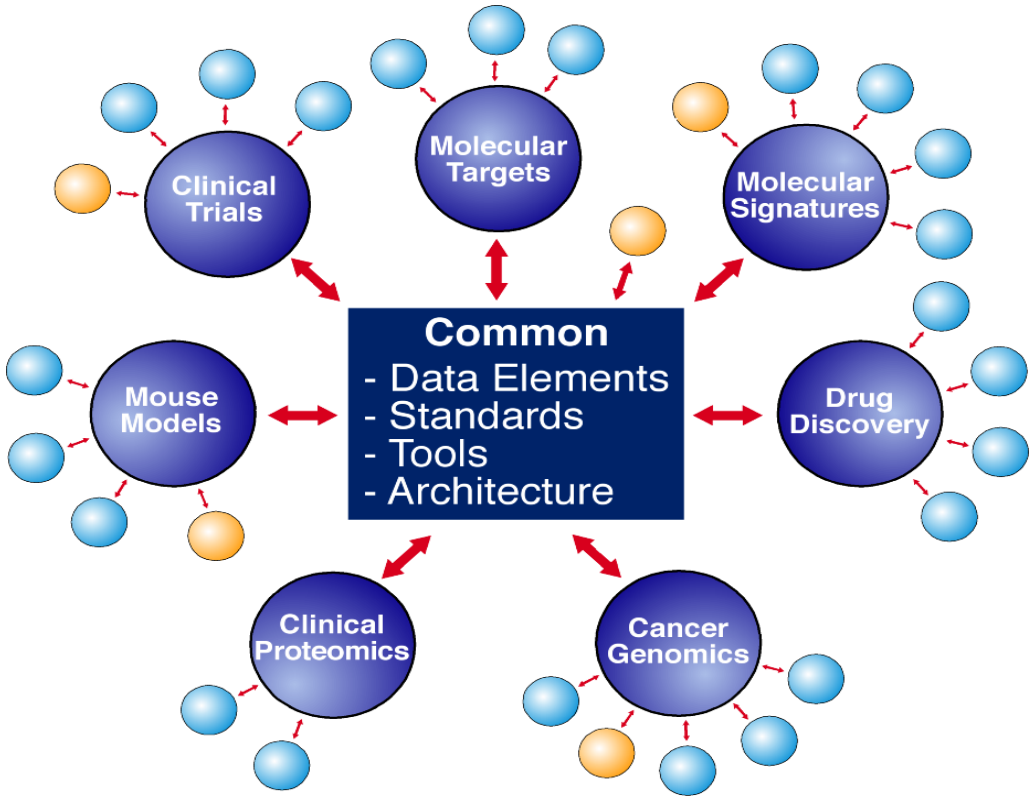



- Common, widely distributed infrastructure permits research community to focus on innovation
- Shared vocabulary, data elements, data models facilitate information exchange
- Collection of interoperable applications developed to common standards
- Raw published cancer research data is available for mining and integration

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# caBIG™'s Informatics Core



# This isn't Rocket Science

- A lot of caBIG™ isn't even computer science
  - Most industries did much of this years ago
- But it is hard to achieve – it takes time
- caBIG™'s goal (oversimplified): facilitate the exchange of data useful for cancer research and care
  - Between research domains, systems, investigators, and organizations
- For instance, the caBIG™ compatibility of a system is determined by how easily the system can exchange data (i.e., interoperability)

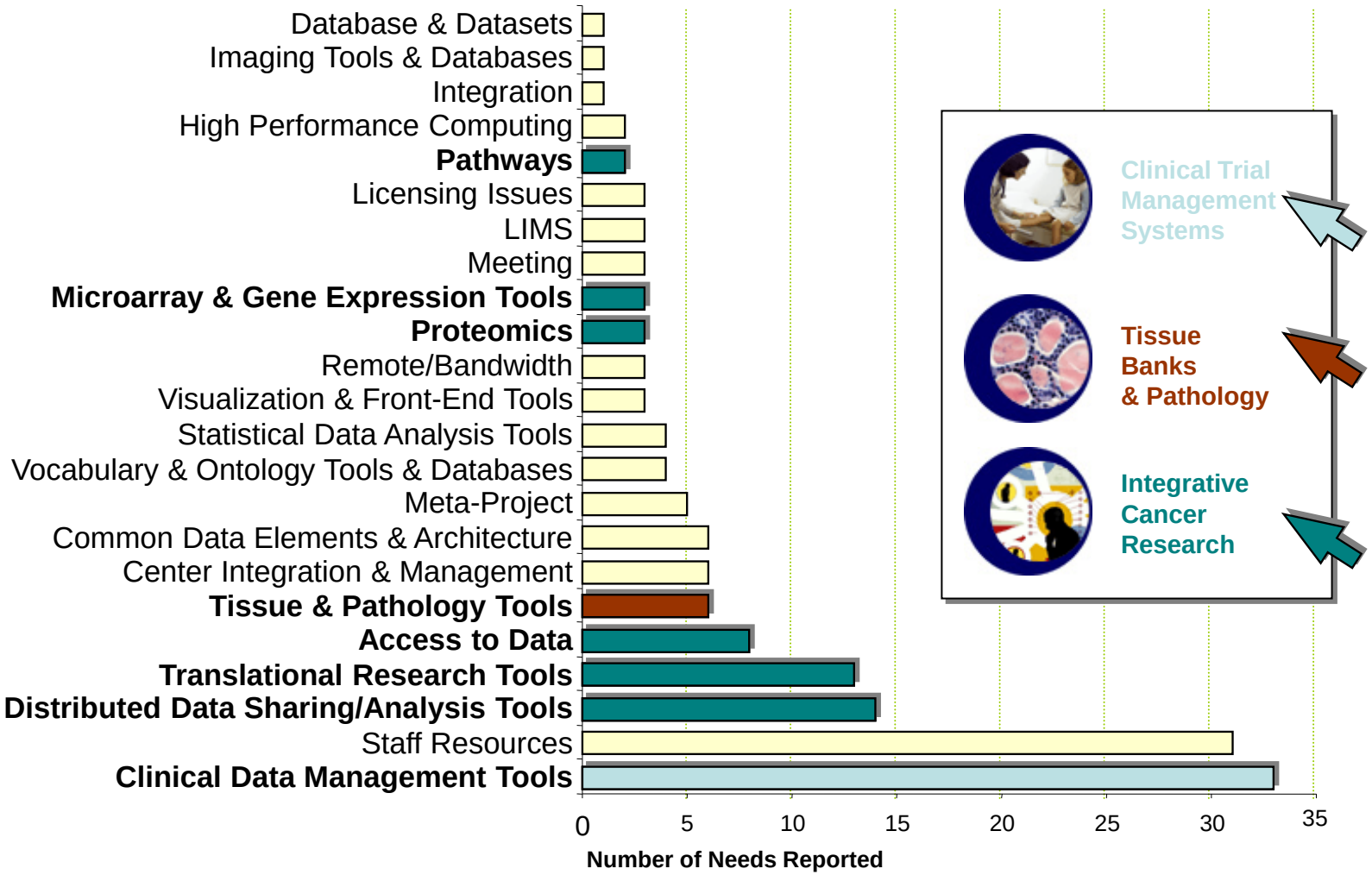
# Talk Agenda

- Introduction and Background
- Program Motivations
- **Program Structure and Process**
- Program Timeline and Products
- Finding Your Way Around caBIG™

National Cancer Institute

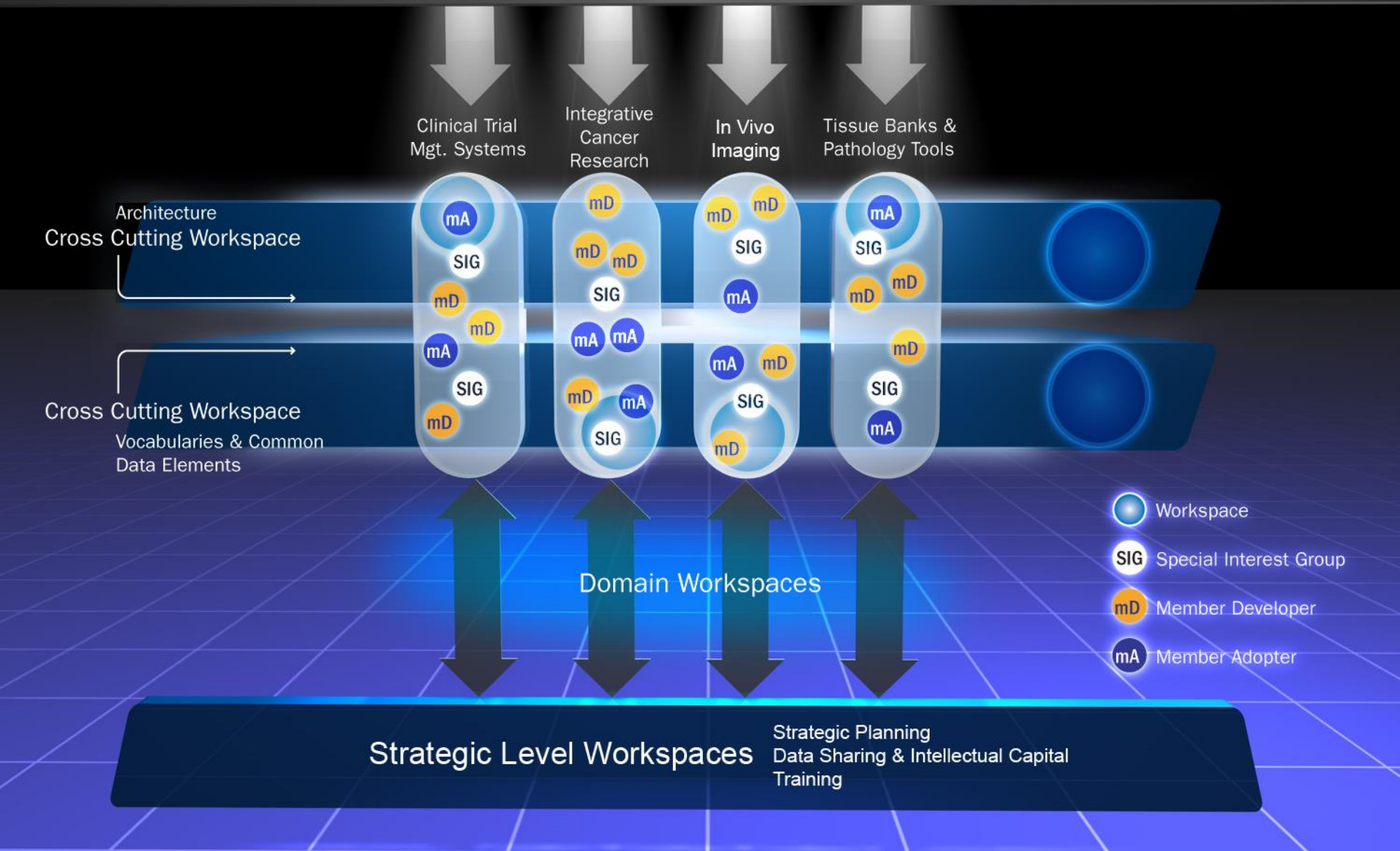
caBIG™ cancer Biomedical Informatics Grid™

# Common needs helped shape priority areas for the caBIG pilot activities



# NCICB

## caBIG™ Program Management Team
## General Contractor



National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# Four Domain Workspaces and two Cross Cutting Workspaces have been launched

**DOMAIN WORKSPACE 1**
**Clinical Trial Management Systems**

Addresses the need for consistent, open and comprehensive tools for clinical trials management.

**DOMAIN WORKSPACE 2**
**Integrative Cancer Research**

Provides tools and systems to enable integration and sharing of information.

**DOMAIN WORKSPACE 3**
**Tissue Banks & Pathology Tools**

Provides for the integration, development, and implementation of tissue and pathology tools.

**DOMAIN WORKSPACE 4**
**Imaging**

Provides for the sharing and analysis of in vivo imaging data.

**CROSS CUTTING WORKSPACE 1**
**Vocabularies & Common Data Elements**

Responsible for evaluating, developing, and integrating systems for vocabulary and ontology content, standards, and software systems for content delivery.

**CROSS CUTTING WORKSPACE 2**
**Architecture**

Developing architectural standards and architecture necessary for other workspaces.

NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# Strategic Level Workspaces

**Data Sharing and Intellectual Capital**

Addresses issues related to the sharing of data, applications and infrastructure both within the consortium and in the larger cancer research community.

**Training**

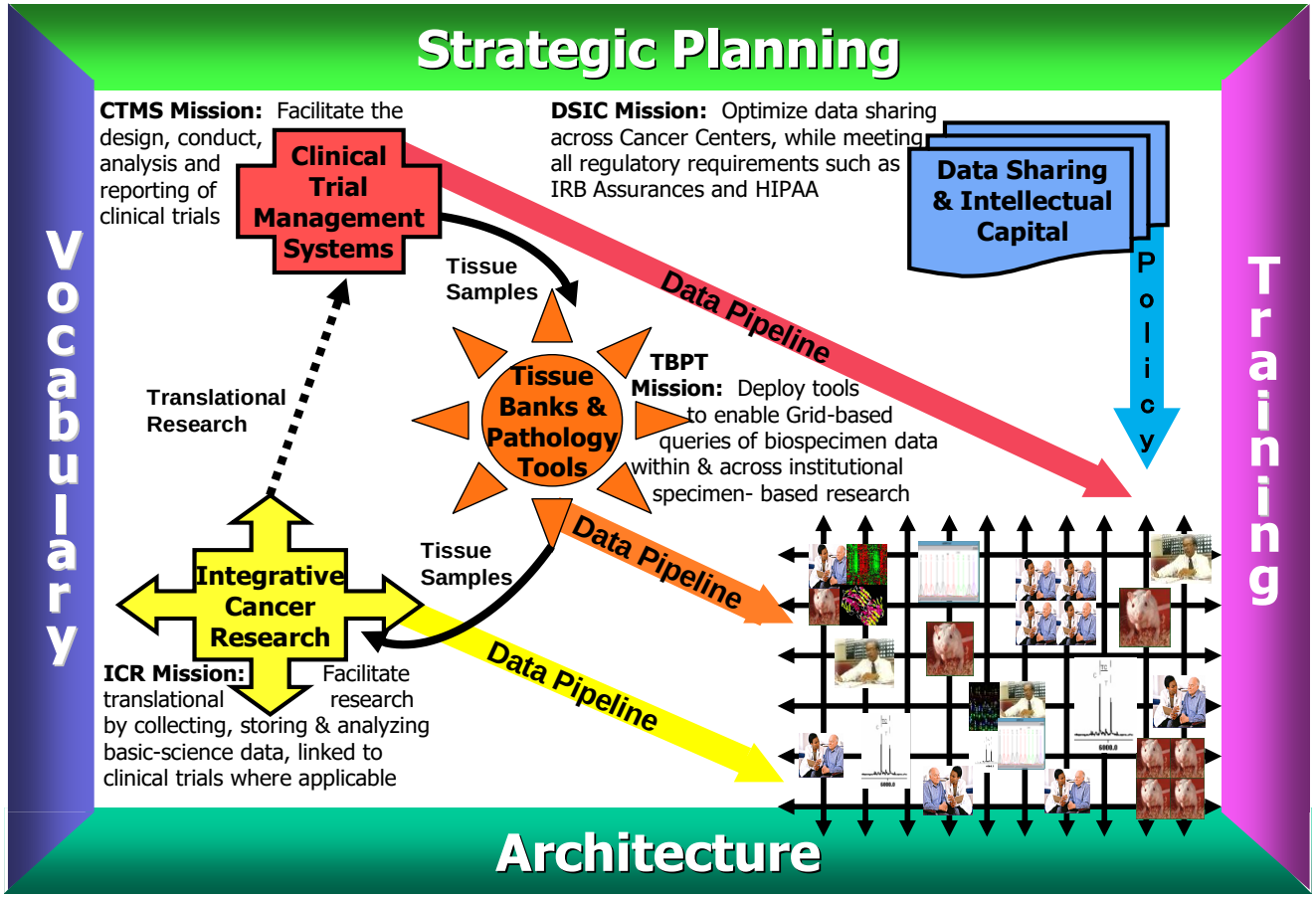
Developing strategies for providing training in the use of the caBIG developed resources including on-line tutorials, workshops, and training programs.

**Strategic Planning**

Assists in identifying strategic priorities for the development and evolution of the caBIG™ effort.

# A Focus on Interaction

Schematic of caBIG Interactions



# caBIG™ principles

- Open source
- Open access
- Open development
- Federated

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# caBIG™: a new way of doing business…

- **Coordinated development**
- **Active management**
- **Community directed**
- **Common services**

# caBIG™ action plan

- **Establish pilot network of NCI Cancer Centers**
  - **Groups agreeing to caBIG™ principles**
  - **Mixture of capabilities**
  - **Mixture of contributions**
- **Expand collection of participants**
- **Establish consortium development process**
  - **Collecting and sharing expertise**
  - **Identifying and prioritizing community needs**
  - **Expanding development efforts**

# Current caBIG™ community

- NCI-designated Cancer Centers (50)
  - Academic Centers (integrated into broader biomedical infrastructure)
  - Stand-alone (community leaders)
  - Community outreach
- NCI Divisions
- NIH NECTAR grantees
- Government
- Industry
- International Groups
  - Standards development organizations
  - U.K.'s National Cancer Research Institute
- ~800 active participants

# Talk Agenda

- Introduction and Background
- Program Motivations
- Program Structure and Process
- **Program Timeline and Products**
- Finding Your Way Around caBIG™

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# Overall Goals for caBIG™

## Three-year (mid-2008)

- Develop sufficient research tools and standards to have a positive impact on the cancer research community, as measured by adoption of relevant caBIG™ principles in project proposals.
- Ensure widespread adoption of standards so that funded projects are operating under the Gold standard of caBIG™ compatibility.
- Adopt and use caBIG™ interoperable tools and data sets within the caBIG™ community.
- Develop mechanisms for engaging and promoting caBIG™ compliant technologies and established datasets within the oncology research community.

NATIONAL INSTITUTES OF HEALTH

NATIONAL CANCER INSTITUTE

# Overall Goals for caBIG™

## Five-year (mid-2010)

- Ensure widespread adoption, dissemination, and use of caBIG™ interoperable tools, standards, and data sets within the larger cancer community, to include the biopharmaceutical industry, non-NCI cancer centers, and the national cancer research enterprise.
- Begin to see results of caBIG™- compliant interdisciplinary and inter-institutional research affecting clinical oncology care.

# caBIG™ Deliverables

- The project is providing useful products in a variety of areas:
  - Clinical Trials (C3D, caAERS, *etc.*)
  - Tissue Banks and Pathology (caTISSUE, caTIES, *etc.*)
  - Bioinformatics (caArray, caWorkbench, *etc.*)
- Tools have been developed by, and with the direct specification of, the Cancer Center community, and reflect the needs and capabilities of that community.

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™




**Building Community ……**

**Delivering Products…....**

**Real Solutions to Real Problems**

NATIONAL INSTITUTES OF HEALTH

NATIONAL CANCER INSTITUTE

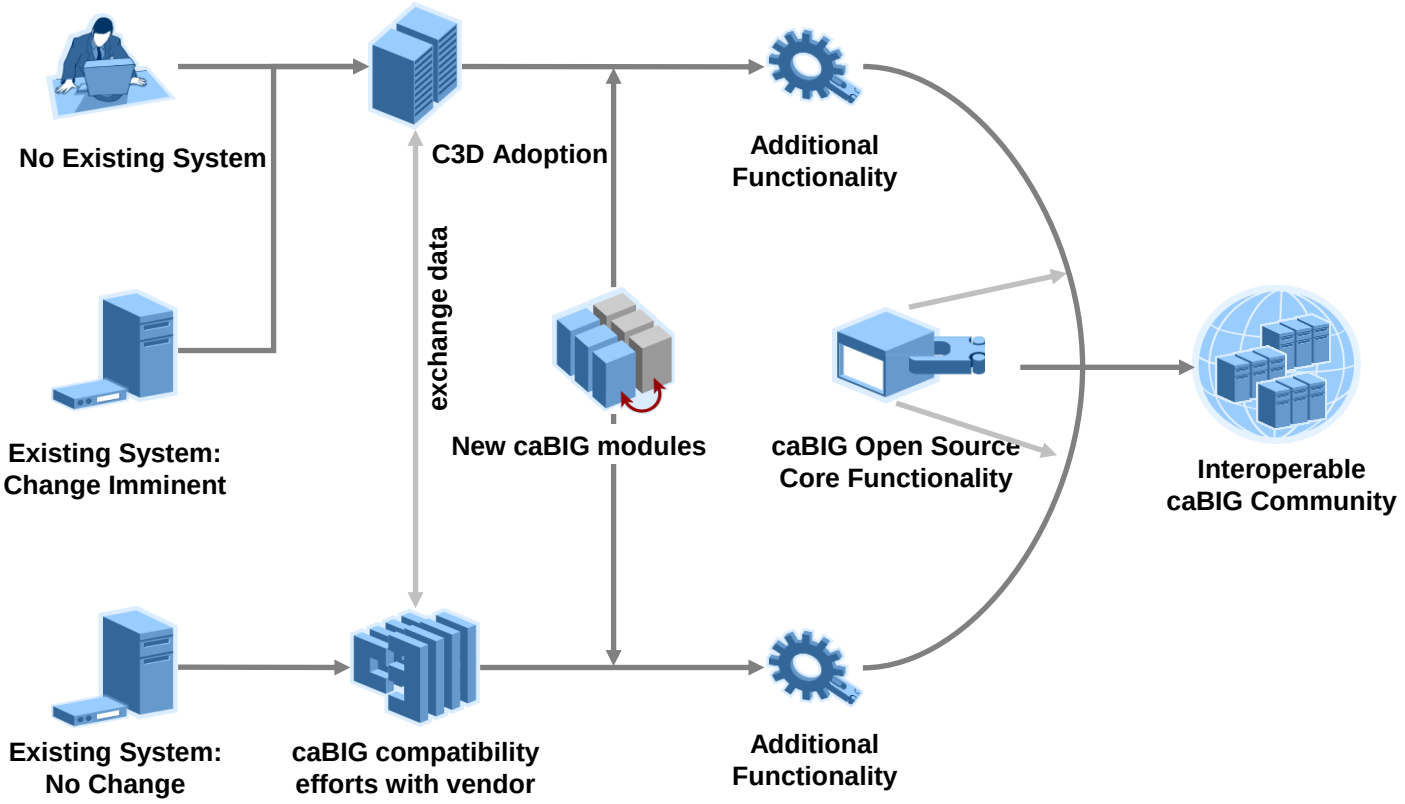

- caBIG™ Activities in …

  Clinical Trials Management Systems


National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# Clinical Trials Management: Three Levels of Interaction Based on Institutional Readiness



National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# 2006 Clinical Trial Tools Development Activities

- caAERS
- Patient Study Calendar
- Lab Data Hub
- Making other CTMS systems caBIG compatible

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# CRIX

## Clinical Research Information eXchange

- Sustainable, secure and standards-based infrastructure for electronic submissions
- Standards-based repository for data analysis and review
- Global registry for commonly used/referenced data
- Common information exchange standards
- Mechanism for secure electronic information exchange
- Legally enforceable digital signatures compliant with Title 21 Regulations and other guidelines from the outset

NATIONAL INSTITUTES OF HEALTH

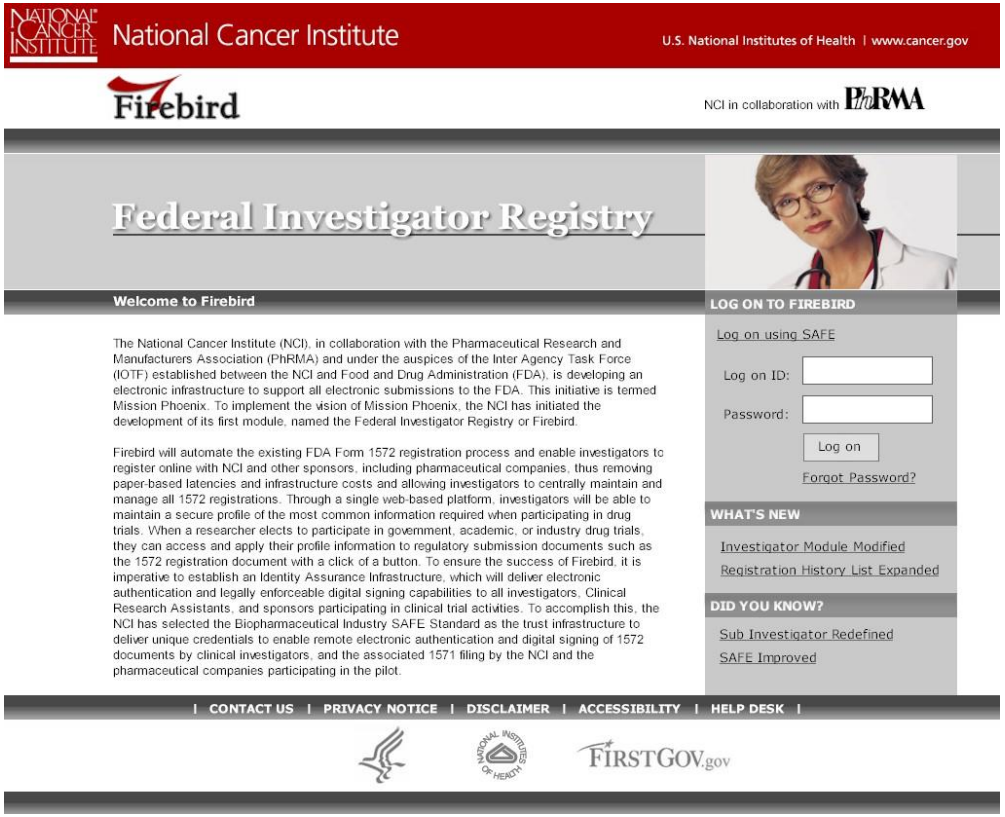
NATIONAL CANCER INSTITUTE

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

## CRIX 1.0: Firebird

- Automates and centralizes the 1572 registration process
- Enables investigators to register online with sponsors
- Proves the feasibility and value of CRIX



# Pilot Project : caMATCH
# Matching Patients to Clinical Trials

- A contribution to strategic research on personal health records and other tools for personal health self-management
- Evaluate the effectiveness of a patient-centric online clinical trials matching program in a pilot project for breast cancer patients in the San Francisco Bay Area

# NCAB Clinical Trials Working Group

- New Initiative: Promote the establishment of national clinical trial information technology infrastructures that are fully interoperable with NCI's cancer Biomedical Informatics Grid.
  - Clinical Trials Working Group of the National Cancer Advisory Board. *Restructuring the National Cancer Clinical Trials Enterprise*, June 2005

# Some directives from the Clinical Trial Working Group Report

- Establish a Comprehensive NCI Clinical Trial Database
- Promote interoperability amongst all clinical trials systems
- Achieve concurrence on standard Case Report Forms
- Develop a credentialing system for investigators and sites that is recognized and accepted by NCI, industry sponsors, clinical investigators, and clinical trial sites

Report located at:
http://integratedtrials.nci.nih.gov/ict/CTWG_report_June2005.pdf


- caBIG™ Activities in …

  Tissue Banks and Pathology Tools

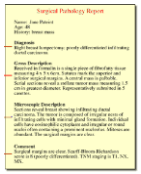
National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# Existing Projects:

## Tissue Banks and Pathology Tools Workspace

- **caTISSUE Core (WU) –** Core specimen handling and tracking functions



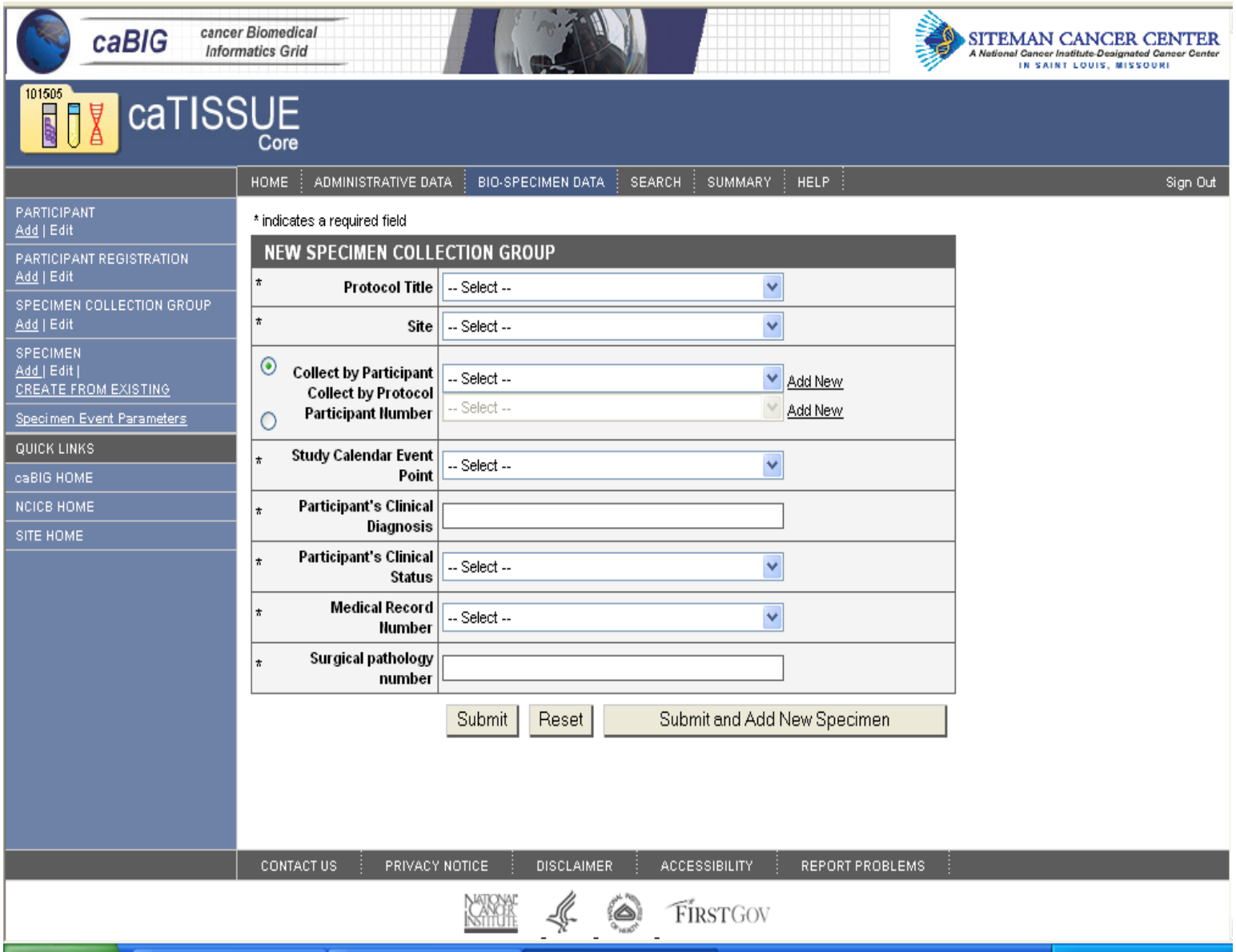
- **caTIES (UPMC) -** Text extraction and de-identification of surgical pathology reports

- **caTISSUE Clinical Annotation Engine (UPMC)** - Annotation of specimens with clinical data

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# caTISSUE Core: Register Specimen Group




- caBIG™ Activities in …

Integrated Cancer Research

# Integrated Cancer Research: Special Interest Groups

- Microarray Repositories
- Data Analysis & Statistics
- Informatics for Proteomics
- Genome Annotation
- Pathways Tools
- Translational Tools
- Population Sciences and Cancer Control

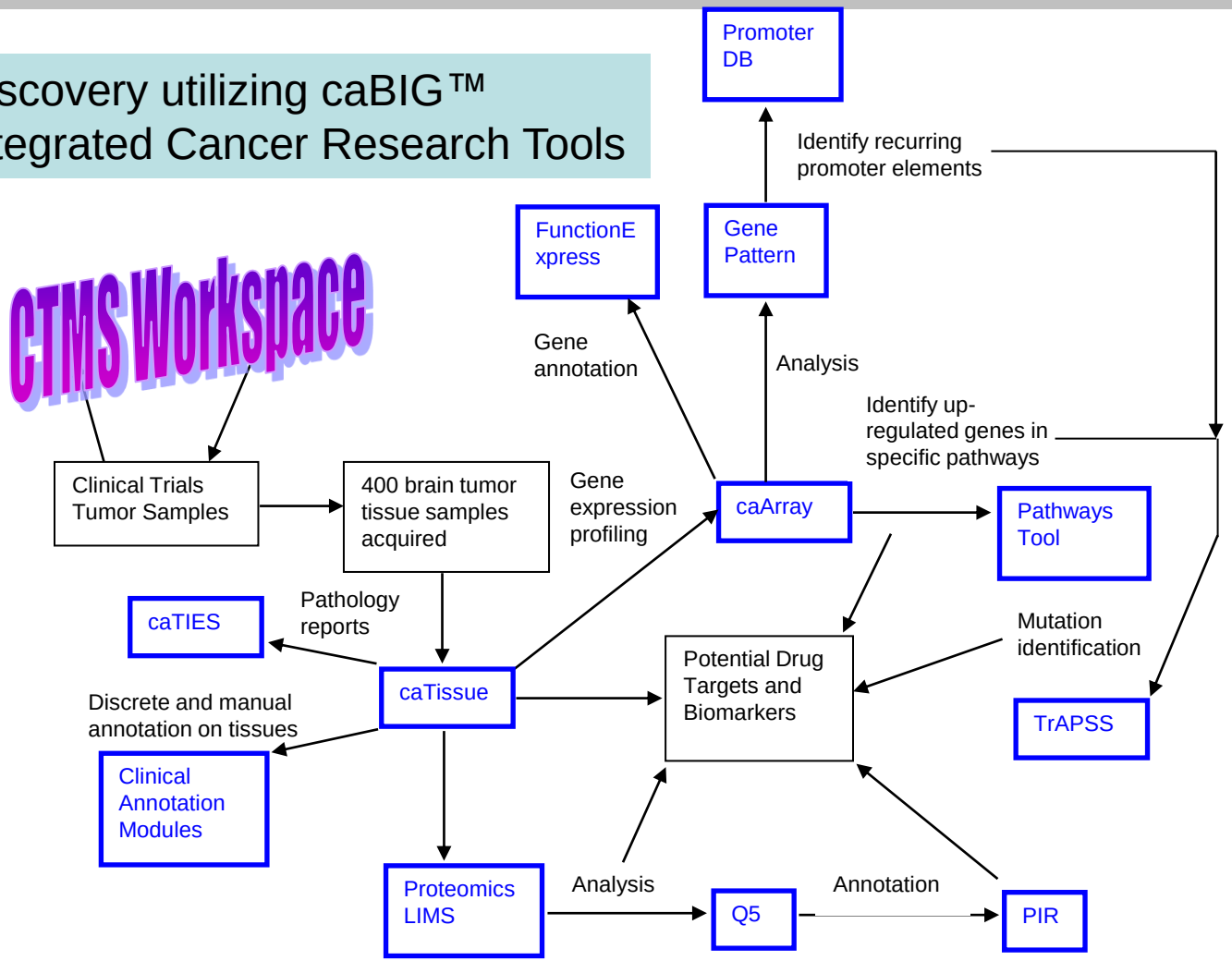
# ICR – New Areas

- Translation Tools
  - RFP out December 2005; proposals due February 6
- Population Sciences and Cancer Control
  - Initial projects March 2006
  - New RFP to follow

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# Discovery utilizing caBIG™ Integrated Cancer Research Tools



# caWorkbench

- **Jointly analyze microarray expression profiles, sequences, motifs, and transcription factors to identify candidate upstream regulators of a particular transcription factor**
- **Customizable, configurable graphical user interface**
- **Visualization & analytical components can be plugged in**
  - **interoperable based on published interfaces**
- **caBIG™ themes:**
  - **Delivering integrated analysis to the biologist's desktop**

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# REMBRANDT: Building a robust translational research framework for brain tumor studies

**RE**pository of **M**olecular **BRA**in **N**eoplasia **DaTa**

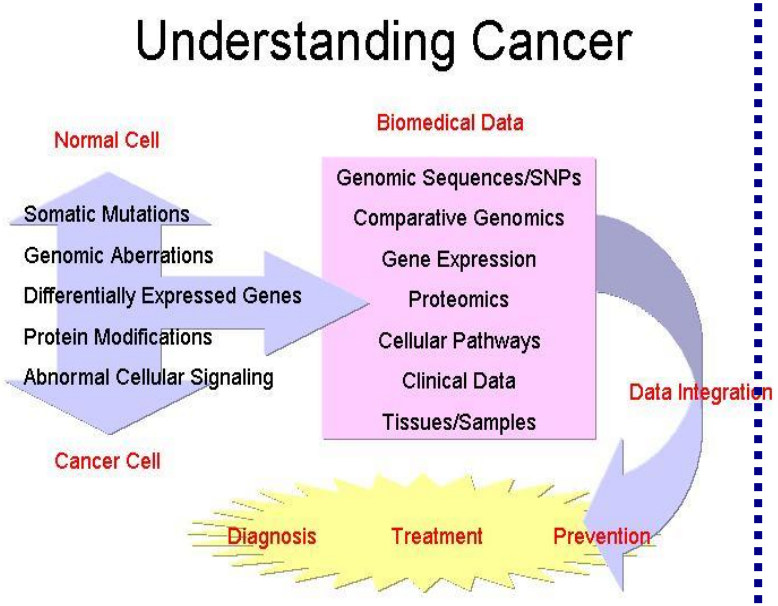
**http://rembrandt.nci.nih.gov**

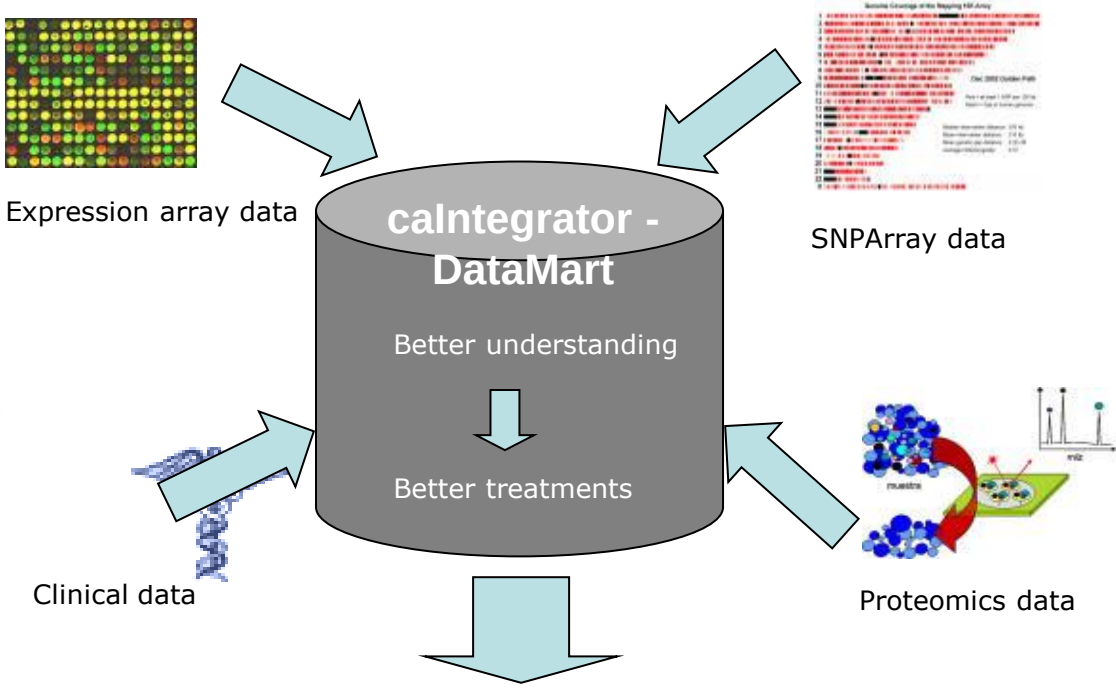
NATIONAL CANCER INSTITUTE · NATIONAL INSTITUTES OF HEALTH · U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES · NCI Neuro-Oncology Branch · National Institute of Neurological Disorders and Stroke · Center for Bioinformatics

nstitute

caBIG™ cancer Biomedical Informatics Grid™

# Rembrandt Knowledgebase

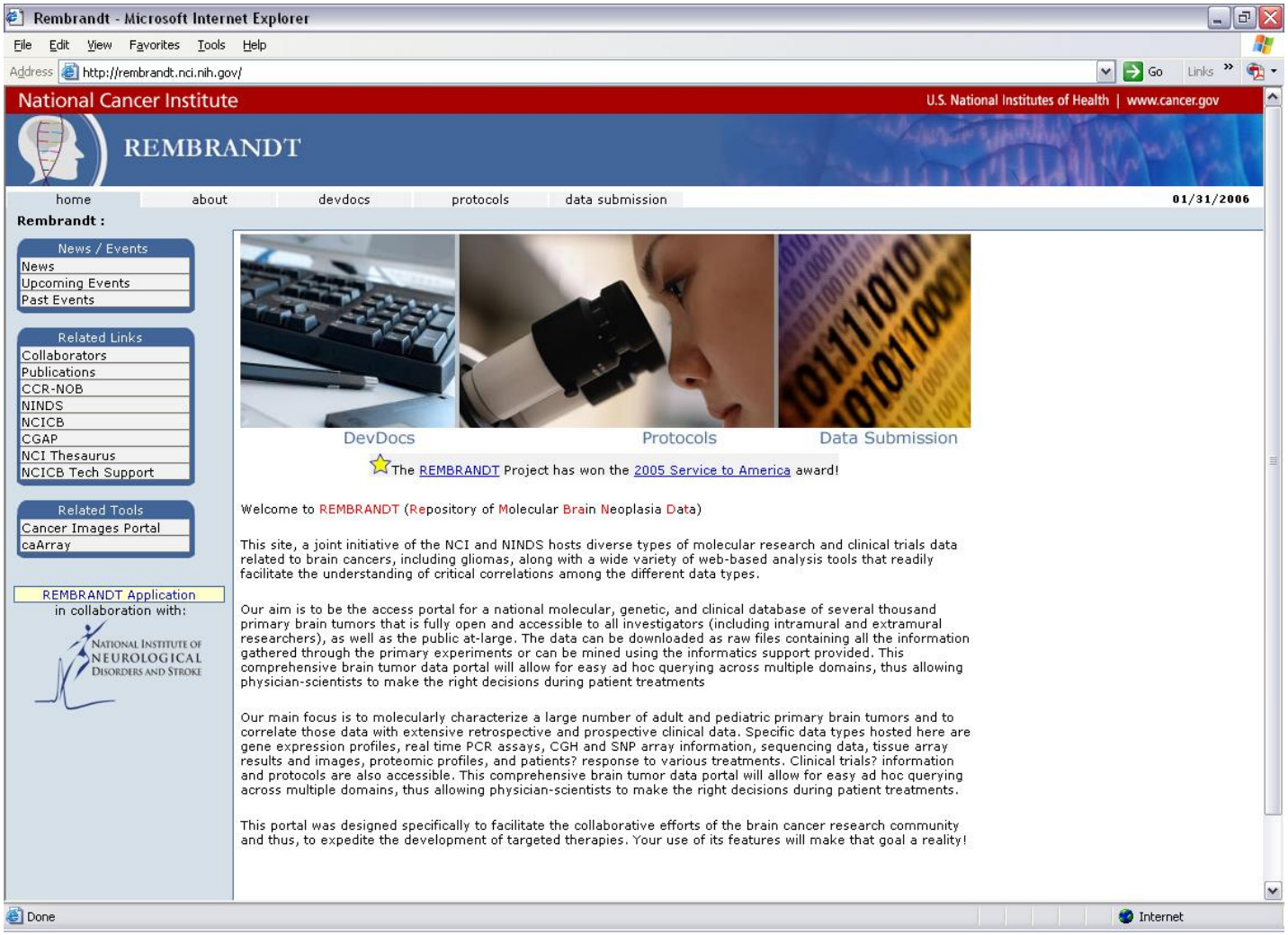



caBIG Analytic Tools

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™




National Institutes of Health | National Cancer Institute

- caBIG™ Activities in …

  In Vivo Imaging

National Cancer Institute

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# caIMAGE – Cancer Images Database

- caIMAGE allows researchers to submit and retrieve images and annotations.
- Images are streamed for efficient access.
- Researchers can search images based on tissue and diagnosis and experiment information.
- Use of common terminology originating from the NCI Enterprise Vocabulary Server (EVS).


NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

- caBIG™ Activities in …

  Vocabularies and Common Data Elements and Architecture

It's all about interoperability and compatibility

# caBIG™ Compatibility

- **caBIG™ is all about Interoperability**
  - Key is to create tools for sharing information
- **Extensible infrastructure**
  - Expandable and modular software to plug into existing systems so current development efforts are not wasted
- **Ensures partnerships**
  - Encourages relationships between academic, government and industry
- **Evolving**
  - Compatibility guidelines are being translated into certification procedures
- **Compatibility Guidelines at https://cabig.nci.nih.gov/guidelines_documentation**

National Cancer Institute

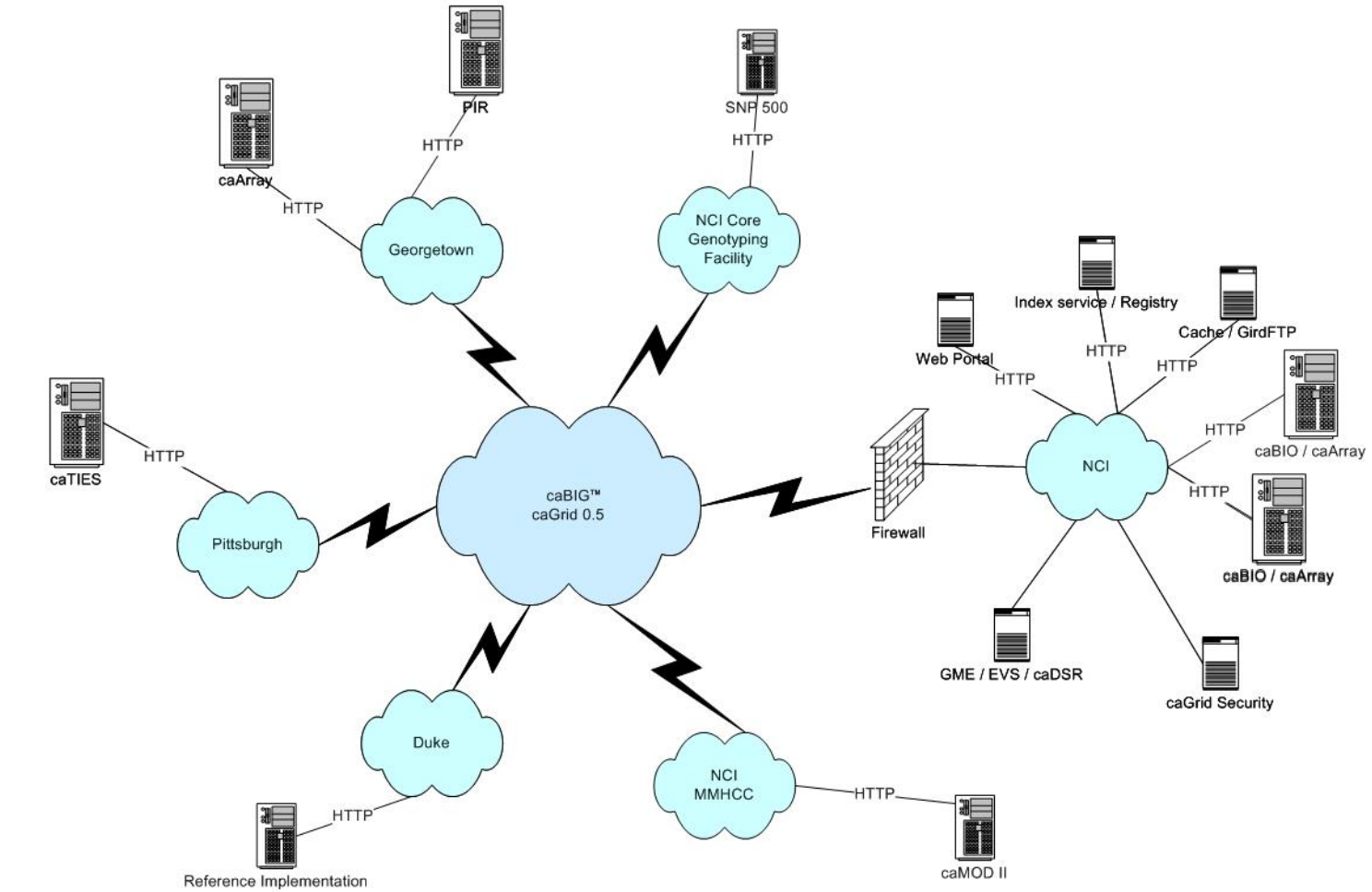
caBIG™ cancer Biomedical Informatics Grid™

# caBIG Compatibility Guidelines

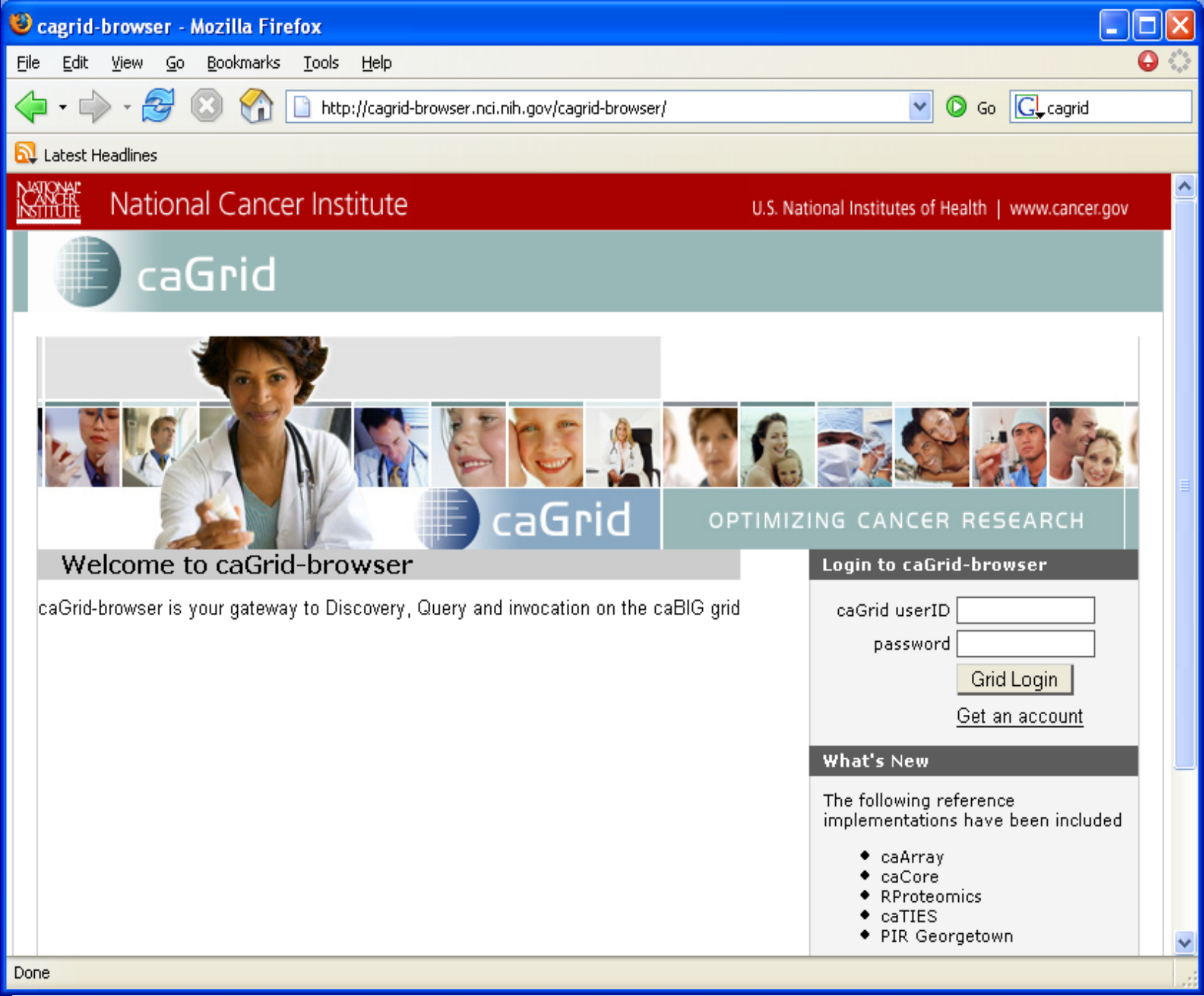
| Maturity Model | Legacy | Bronze | Silver | Gold |
|---|---|---|---|---|
| Interface Integration | - No Programming interfaces to the system are available. Only local data files in a custom format can be read<br>- Some ad hoc data transfer mechanism such as FTP | - Provide baseline* programmatic access to data. Data can be read from remote electronic sources or from commonly used file formats Data can be pushed out to from applications to other external data sources | - Well-described API's that provide access to data objects.<br>- System architecture separated into tiers and interoperable components<br>- Data read in from standards-based electronic sources that support standard or commonly used interchange formats<br>- Documented component description of the underlying data structures that are accessible<br>- Standard messaging systems where appropriate | - All features of Silver, plus:<br>- Interoperable with data grid architecture to be defined by caBIG<br>- Fully componentized provide access to individual resources in the form of grid services |
| Vocabularies / Terminologies & Ontologies | - Free text used throughout for data collection | - Use of publicly accessible standardized controlled vocabularies as well as local terminologies | - Standard terminologies approved by public standards bodies or the caBIG Vocabulary/CDE Workspace are used for all relevant data collection fields. | - All features of Silver, plus:<br>- Fully compliant with caBIG recommended standards for vocabulary terminology services and content sources |
| Data Elements | - No Structured metadata is recorded | - Some type of metadata describing the information in the system is used for data collection and external reporting. Metadata is retrieved from external repository shared by multiple applications.<br>- Common Data Elements should be built using controlled terminology | - Use common standard electronic representation for CDE's such as ISO 11179 or comparable standard<br>- CDEs are harmonized and re-used from across the Domain Workspace<br>- Common Data Elements are built using standard controlled terminologies approved by public standards bodies or the caBIG Vocabulary/CDE Workspace | - All features of Silver, plus:<br>- Programmatic access to all metadata, including data class descriptions, site and source information, and any other caBIG-defined metadata requirements and use information models<br>- Use the caBIG standard or electronic representation of metadata and Common Data Elements |
| Information Models | - No particular information model is used to represent data | - Some type of diagrammatic model describing the data relationship is available in electronic format | - Information models defined in a standard modeling language such as UML | - All features of Silver, plus:<br>- Information models are harmonized with other s across the caBIG Domain Workspace |

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™



National Cancer Institute



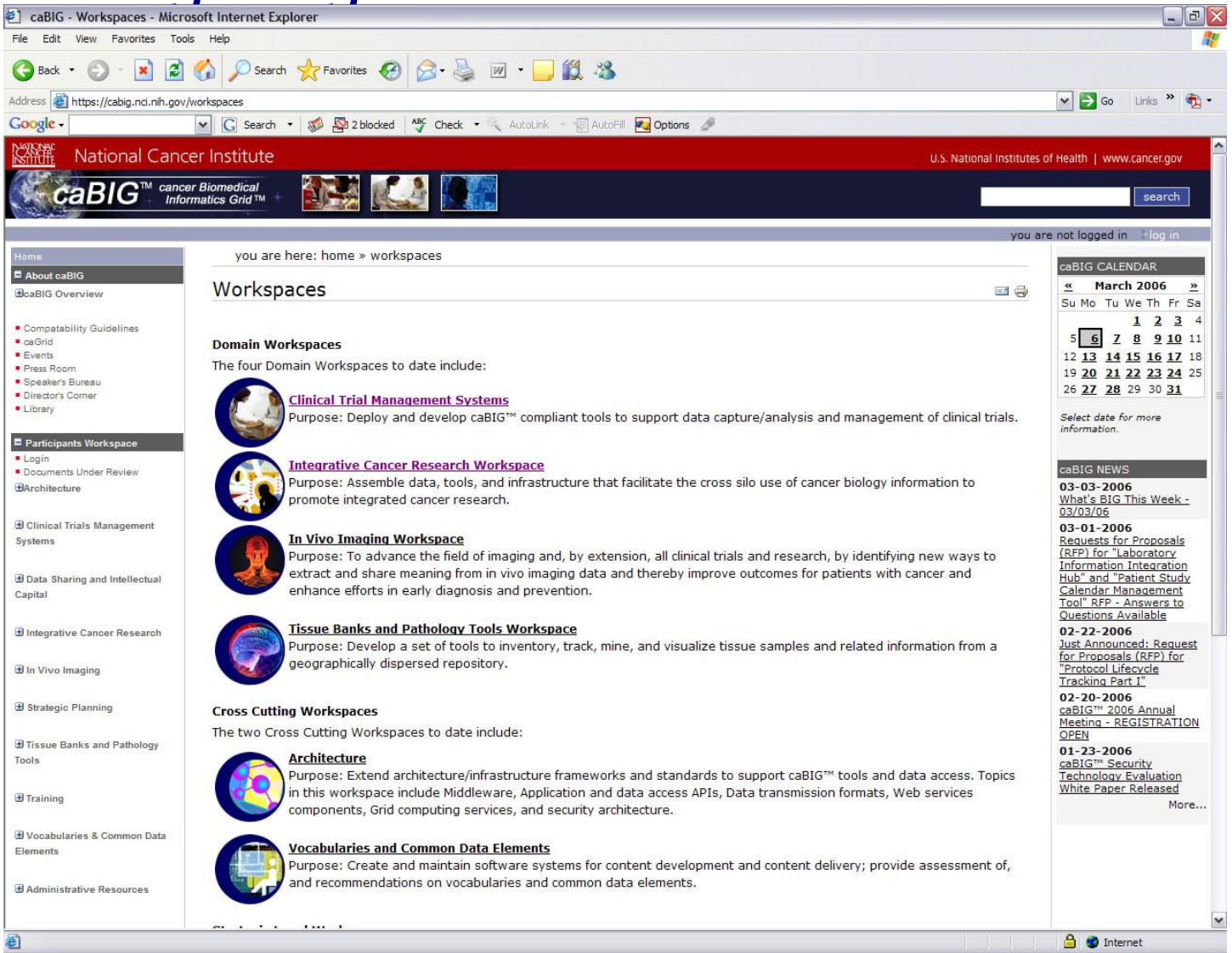
# Talk Agenda

- **Introduction and Background**
- **Program Motivations**
- **Program Structure and Process**
- **Program Timeline and Products**
- **Finding Your Way Around caBIG™**

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

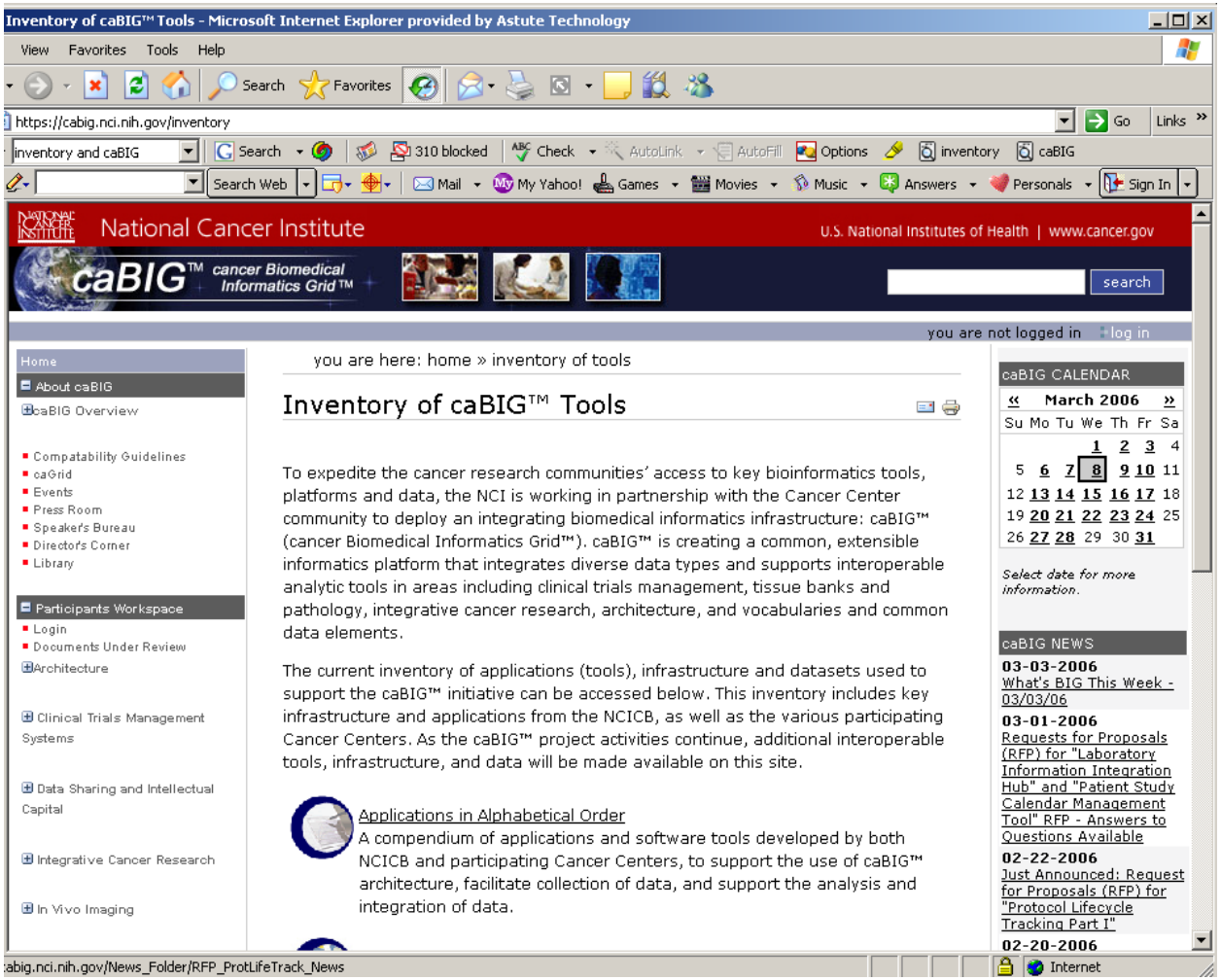
# Navigating the caBIG™ Website - 1



National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# Navigating the caBIG™ Website - 2



# List of Tools

- caBIG™Program Update March 2006
- This issue spotlights caBIG™ products currently available and pending release in 2006 - 2007, and highlights the release of caGrid Software Version 0.5
- cabig.nci.nih.gov/Program_Updates/cabig_March_2006_Program_Update.pdf

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# How can my research benefit from caBIG™ Tools?

- **Everything developed by the program is open source and freely available**
- **Training is available at https://cabig.nci.nih.gov/training**
- **The latest versions of all the software developed as part of the project can be obtained from the caBIG™ CVS site:**
  - **http://cabigcvs.nci.nih.gov/viewcvs/viewcvs.cgi/**
- **Commercial-grade documentation is provided as part of the project, which will be located at the project gforge site:**
  - **http://gforge.nci.nih.gov**

NATIONAL INSTITUTES OF HEALTH NATIONAL CANCER INSTITUTE

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# How can I get support for these tools?

**NCICB Applications Support will coordinate support for caBIG™ tools:**

- Live Support: Monday – Friday 8 am – 8 pm Eastern Time
  - Telephone support is available Monday to Friday, 8 am – 8 pm Eastern Time, excluding government holidays.
  - You may leave a message, send an email or submit a support request via the Web at any time.
- Email: **ncicb@pop.nci.nih.gov**
- Phone: **301-451-4384**
- Toll-free: **888-478-4423**
- Web: **http://ncicbsupport.nci.nih.gov**

National Institutes of Health

National Cancer Institute

# caBIG™: Getting Involved

- To get involved with caBIG™:
  - Track caBIG™ activities on the NCI's caBIG™ website, https://cabig.nci.nih.gov/
  - Attend caBIG™ Annual Meeting, April 9-11, 2006, Hyatt Regency Crystal City, Arlington, Virginia
  - Learn about the existing bioinformatics infrastructure, caCORE, at https://ncicb.nci.nih.gov/core
  - Download currently available caBIG™ tools from the caBIG™ website at https://cabig.nci.nih.gov/inventory
  - Sign up for the caBIG™ mailing list at http://list.nih.gov/archives/cabig_announce.html
- Please visit the main caBIG™ website for more information: https://cabig.nci.nih.gov/

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# Save the Date!

- **The caBIG™ 2006 Annual Meeting**
- **April 9-11, 2006**
- **Hyatt Regency Crystal City, Arlington, Virginia**
- **Plenary sessions; 35 break out sessions; dozens of demonstrations, and posters; exhibits; hackathon**
- **Tailored sessions for newcomers April 9 and throughout the conference**
- **https://cabig.nci.nih.gov/2006_Annual_Meeting**

NATIONAL INSTITUTES OF HEALTH

NATIONAL CANCER INSTITUTE

National Cancer Institute

caBIG™ cancer Biomedical Informatics Grid™

# Contact Information

Mary Jo Deering, Ph.D

Director for Informatics Dissemination

NCI Center for Bioinformatics

National Cancer Institute

National Institutes of Health, USDHHS

6116 Executive Blvd. - #403

Rockville, MD  20852

(o) 301-496-3458

(f) 301-480-4222

deeringm@mail.nih.gov