National Cancer Institute

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

caBIG®: A Translatable Informatics Platform for 21st Century Biomedicine

Annual Federal Networks Conference

February 15, 2011

Ken Buetow, Ph.D.

Director, Center for Bioinformatics and Information Technology

National Cancer Institute

The Cancer Community Seeks To*…

- **Manage/Leverage local and publicly-accessible biomedical data for research**
- **Connect/Streamline workflows**
- **Increase accuracy of processes**
- **Standardize/Streamline data collection**
- **Perform complex analysis across data sets**
- **Identify best practices**
- **Share data**

** Based on 2004 Cancer Centers survey*

caBIG® - Building a Worldwide Web of Cancer Research: The Government Role 2003-2010

caBIG® Strategy at Launch

- **Community**
  - Establish an **open community** of participants from the spectrum of disciplines, geographies, types of institutions, etc.
  - Facilitate the work of others who are building capabilities
  - Adopt a **"federated"** model to allow local control of sharing and partnerships and to support individual labs and institutions
- **Content**
  - Facilitate access to rich primary data
  - Leverage existing academic and commercial software, wherever possible, to avoid unnecessary time and expense
  - Invest primarily in **open source** tools that the community does not have
- **Connectivity**
  - Recognize legacy IT systems to avoid "rip and replace" costs
  - Wherever feasible, make disparate applications compatible for "plug-and-play" compatibility and data-sharing through **standards-based interoperable infrastructure**

caBIG® Is a Platform to Link All Major Functions in the R&D Process

caBIG® Is a Platform to Link All Major Functions in the R&D Process

- Track clinical trial registrations
- Facilitate automatic capture of clinical laboratory data
- Manage reports describing adverse events during clinical trials

Clinical Research

Imaging

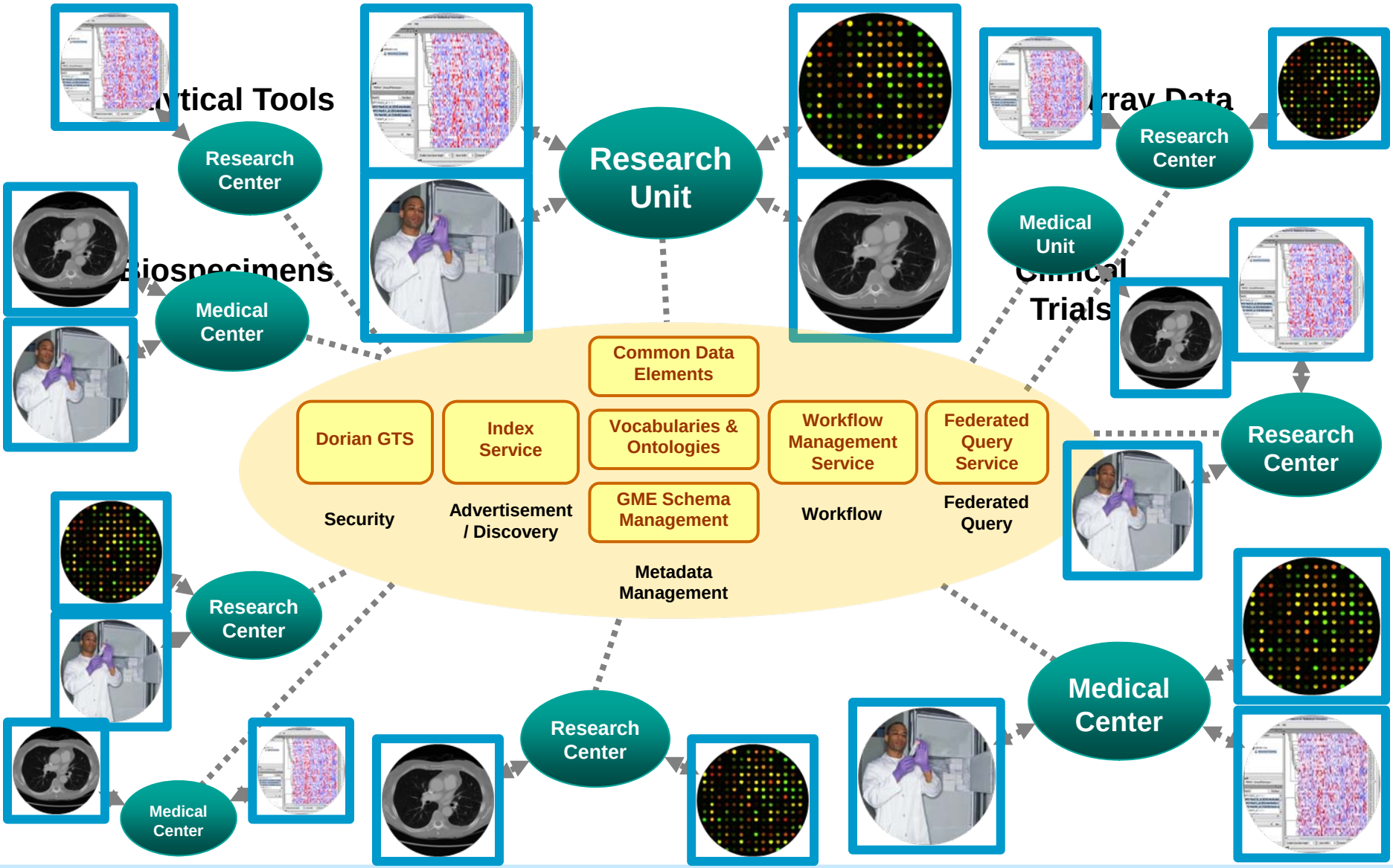
- Utilize the National Cancer Imaging Archive repository for medical images including CAT scans and MRIs
- Visualize images using DICOM-compliant tools
- Annotated Images with distributed tools

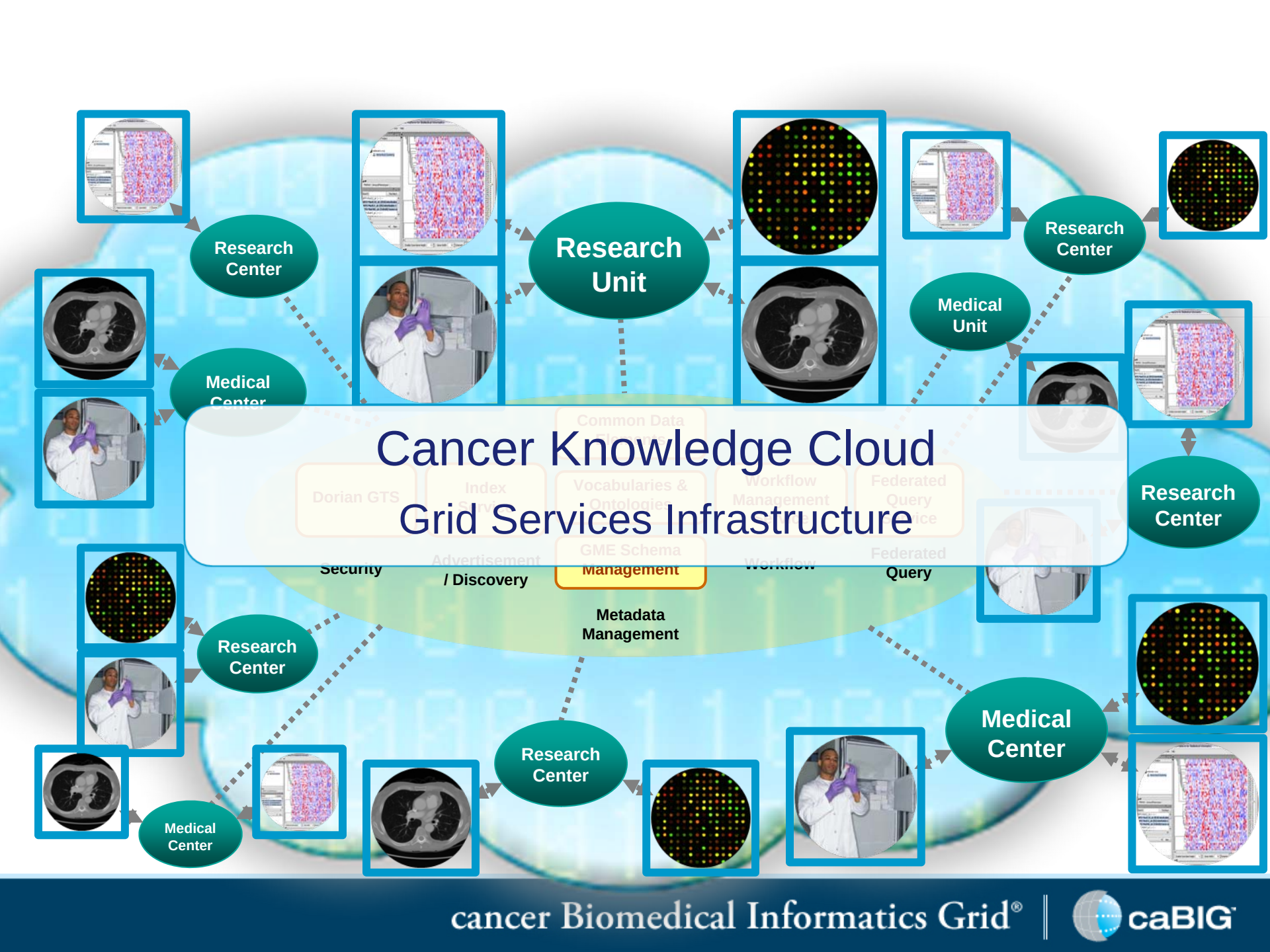
- Combine proteomics, gene expression, and other basic research data
- Submit and annotate microarray data
- Integrate microarray data from multiple manufacturers and permit analysis and visualization of data

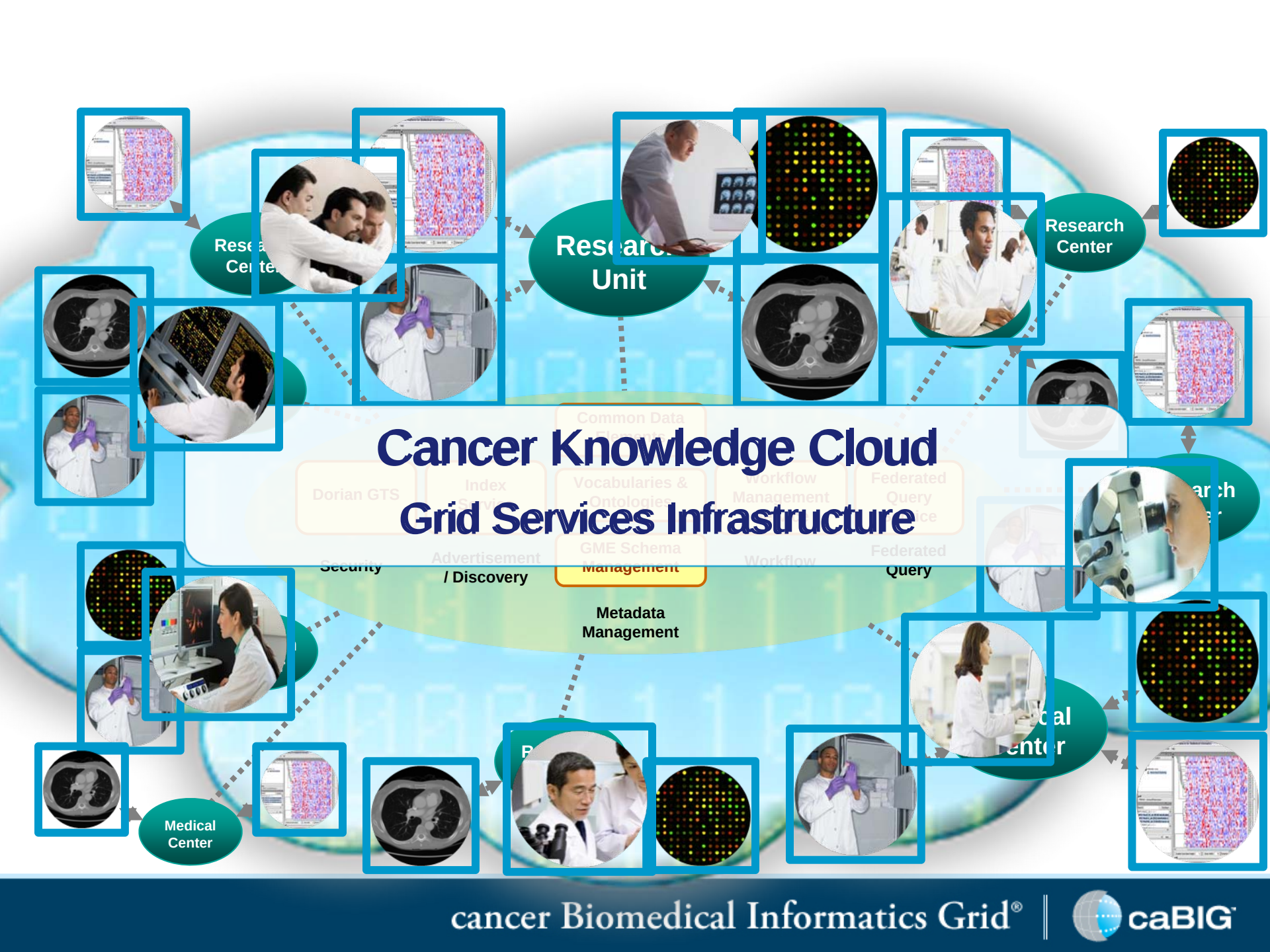
Molecular Biology

Pathology

- Access a library of well characterized, clinically annotated biospecimens
- Use tools to keep an inventory of a user's own samples
- Track the storage, distribution, and quality assurance of specimens


ytical Tools
Research Center
Research Unit
rray Data
Research Center
Biospecimens
Medical Center
Medical Unit
Trials
Common Data Elements
Dorian GTS
Index Service
Vocabularies & Ontologies
Workflow Management Service
Federated Query Service
Research Center
Security
Advertisement / Discovery
GME Schema Management
Workflow
Federated Query
Metadata Management
Research Center
Research Center
Medical Center
Medical Center


Research Center
Research Unit
Research Center
Medical Unit
Medical Center
Research Center
Cancer Knowledge Cloud
Grid Services Infrastructure
Dorian GTS
Index Service
Vocabularies & Ontologies
Workflow Management
Federated Query
Security
Advertisement / Discovery
GME Schema Management
Workflow
Federated Query
Metadata Management
Research Center
Research Center
Medical Center
Medical Center
cancer Biomedical Informatics Grid®
caBIG



Cancer Knowledge Cloud

Grid Services Infrastructure

caBIG® Today

Community

- 2,300+ participants from more than 700 institutions
  - 56 NCI-designated Centers
  - 30 Community Centers
- 1000+ registrants for the 2010 caBIG® Annual Meeting
- 19 licensed Support Service Providers to sustain the biomedical community as they deploy caBIG® tools and technology
- 15 countries using caBIG® tools and technology to facilitate

Connectivity

- 78 applications supporting full continuum of biomedical research
- 149 “nodes” connected to National Grid via caGrid

Content

- 2.17 million biospecimens available through caGrid
- 4.76 million images stored in the National Biomedical Imaging Archive
- 39,952 microarray experiments available for research use on caGrid

caBIG® Is Expanding Internationally To Become a Global Network

15 countries engaged with and/or using caBIG® tools and technologies, including:

- United Kingdom
- Latin America
- India
- China
- Mexico, Brazil, Uruguay, Argentina, Chile
- Czech Republic
- The Netherlands
- Germany
- Finland
- Jordan
- Pakistan
- Australia
- New Zealand

- **United Kingdom:** NCI collaboration with National Cancer Research Initiative (NCRI) focused on the use of caGrid technology to connect researchers and enable exchange of research data.
- **Latin America:** Latin American Cancer Pilot Program began patient enrollment for two clinical trials on molecularly characterized stage II and III breast cancer patients, using a broad suite of caBIG® tools to ensure that researchers can compare data across partner sites.
- **China:** Duke University Comprehensive Cancer Center and Beijing University Cancer Hospital launched a collaboration using caBIG® capabilities to conduct first clinical trials in China where all patients are registered electronically.
- **India:** NCI engaged with the All India Institute of Medical Sciences, the Center for Development of Advanced Computing, and the Tata Memorial Hospital of Mumbai over the use of grid computing for managing clinical trials data.

"From the political side, oversight committees in Congress – particularly in the House – will be looking for opportunities to hold hearings, write reports, and pressure the administration to limit government spending on programs of record....

This increased oversight from the political and administrative sides will introduce new levels of budget and schedule uncertainty into large IT programs."

Warren Suss, Conference Chairman
Opening Remarks for the 24th Annual Federal Networks Conference

National Cancer Institute

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

2011 and Beyond:

A Public-Private Biomedical Ecosystem Enabled by Translatable Informatics

21st Century Biomedical Research Still Faces Huge Data/Knowledge Challenge

- **21st Century research is based on a continuum – from bench to bedside and back, and makes use of digital "data" of all kinds – biological, molecular, clinical, laboratory, pharmacy, etc. – to drive knowledge.**

- **But….. at present, these data are still:**
  - Of varying quality;
  - Non-conformant to standard vocabularies;
  - Frequently incomprehensible or prohibitively laborious to translate from one discipline to another.

- **Unlike almost all sectors in today's knowledge economy, biomedicine has no means for efficient collection, aggregation, integration, analysis, interpretation, and transmittal of data so that it can be converted into practical, useful knowledge by anyone other than the original author.**

How Translatable Informatics Can Meet the Challenge

- **Translatable informatics (TI) is an open information technology framework – comprised of standards, specifications, vocabularies, and code bases – that enables users to capture, aggregate, integrate, analyze, interpret, and transmit data through open interfaces between different repositories, institutions, or other sources of data.**

- **TI can facilitate the translation of data into information at all points of the biomedical life cycle (discovery science, translational science, clinical development, clinical care, population science.)**

Translatable Informatics generates the flow of information necessary to sustain all participants in the biomedical ecosystem.

Who Needs Translatable Informatics?

- **Pharmaceutical companies need a reusable clinical trials infrastructure to avoid "re-creating" every clinical trial de novo, especially for molecularly targeted drugs in which patient subgroups must be identified early in the process.**
- **Research institutions need a common framework to use data trapped within their own different departments and laboratories, to fuel collaborations and accelerate time to discovery.**
- **Individual researchers need "liquid data' at their fingertips that can be used to form or validate research hypotheses, empowering them to leverage more diverse and/or larger datasets than those they have themselves generated.**
- **Clinical care providers need to mobilize the clinical data already in their systems to measure the efficacy of their current care, to improve clinical outcomes, and to fuel research by correlating clinical profiles with molecular data.**
- **Patient advocacy organizations need cost-effective platforms to gather clinical information from their patient population as well as accompanying research data generated from those populations, to help physicians determine most effective care and to help drive therapy-focused research by academics and pharmaceutical companies.**

The Role of the Commercial IT Sector

- **Software developers and systems integrators need to provide products and services customized to the biomedical market.**

- **Since the use of TI requires only *open interfaces* between systems, it does not disrupt the development and commercialization of proprietary information technology products and services.**

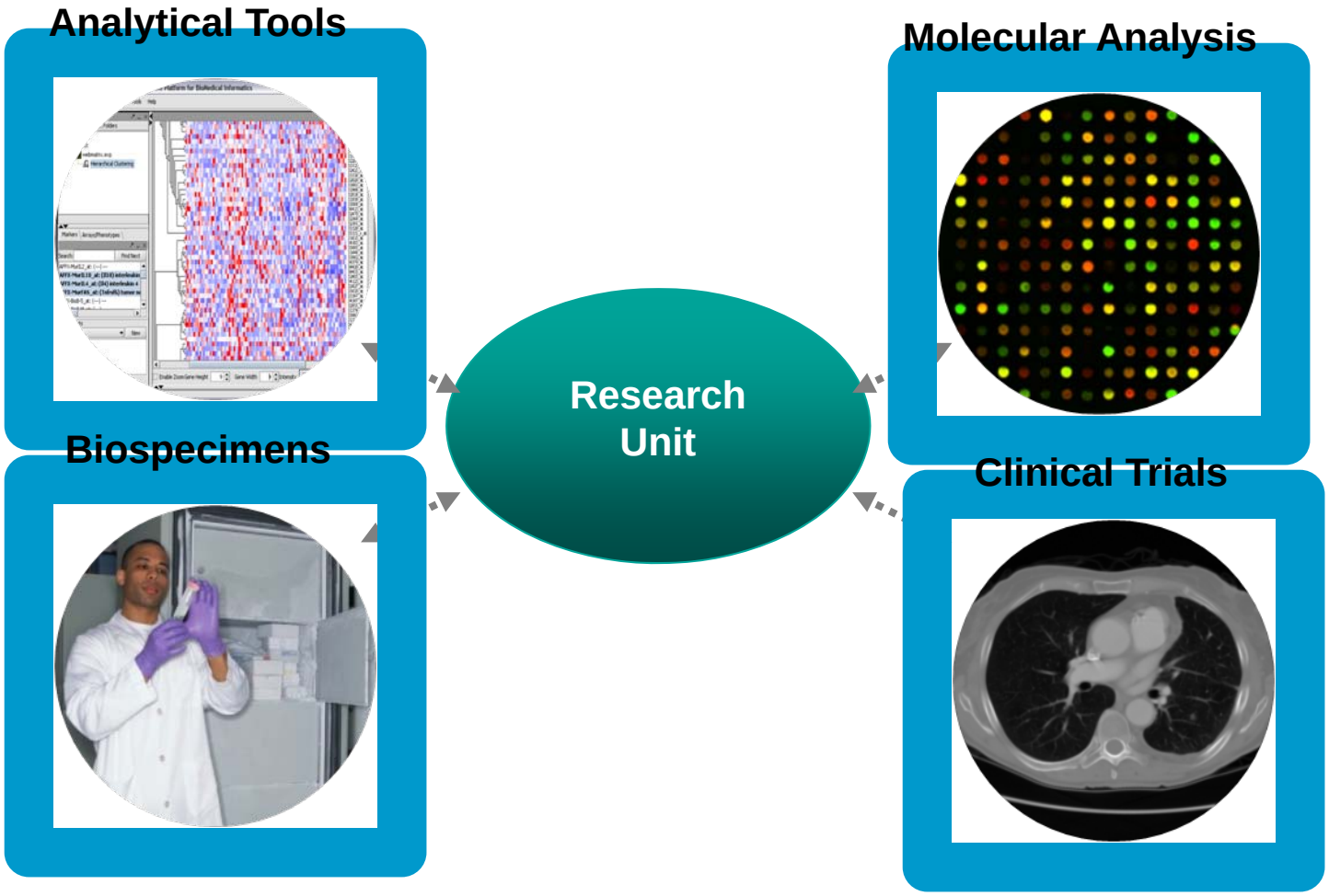
- **Such IT companies will benefit from:**
  - The creation of a “common market” evolved from the fragmented components of the different sectors;
  - The reduction of risk that results from having a core collection of non-viral, open source code that can be reused;
  - The opportunity to leverage a substantial investment already made by the government in caBIG® as a common informatics infrastructure

Requisites for Translatable Informatics

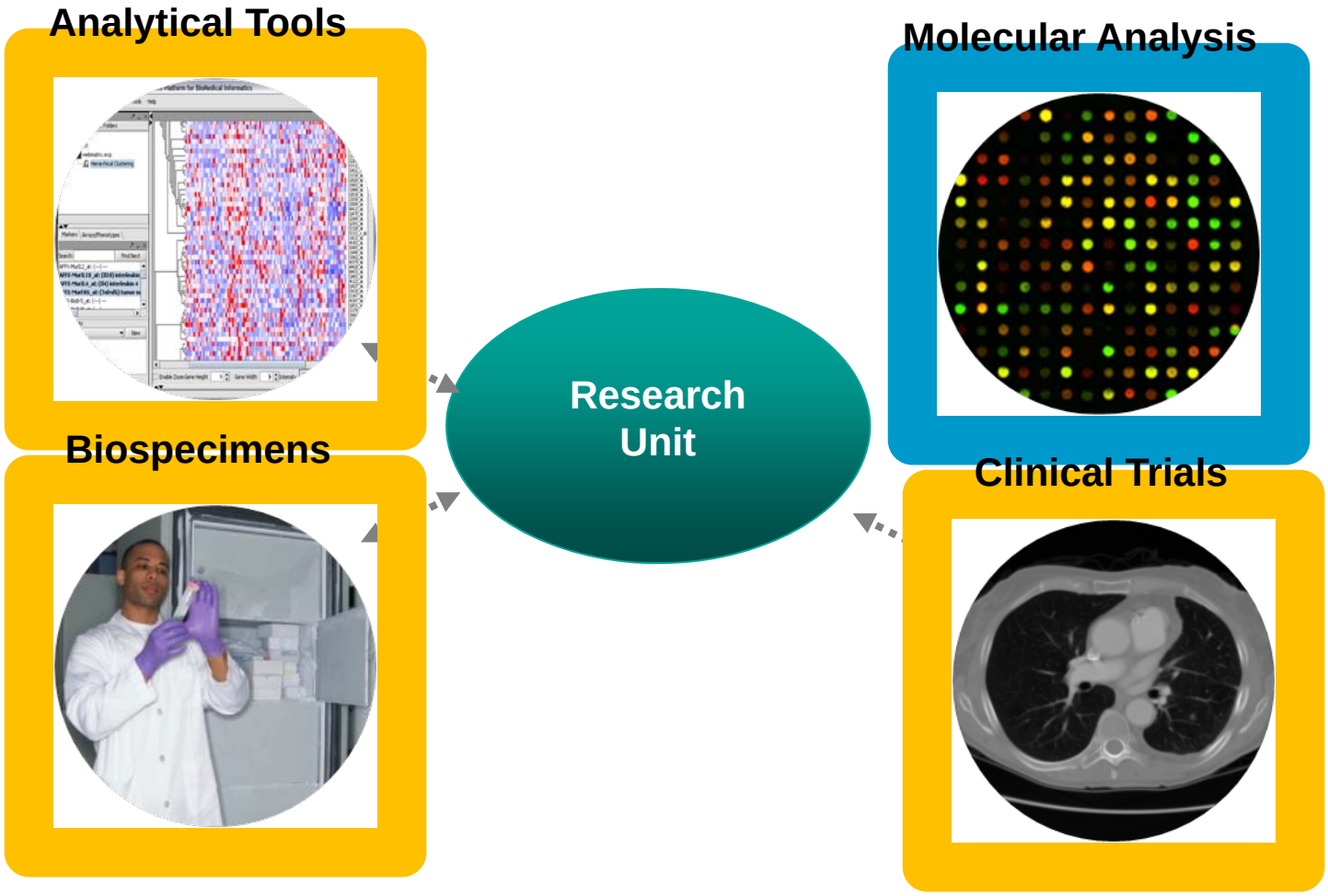
- **Data Standards: Shared data standards ensure that data generated at different locations and by different applications is accessible/meaningful.**
- **Software code: This includes critical capabilities to support the generation and management of basic and clinical research data, management of clinical trials, storage and analysis of medical images, and acquisition, management, and tracking of biospecimens.**
- **Grid or network: A grid that provides the "backbone" to which applications and data sources can connect, and allows researchers to access shared data via the use of web-query tools.**
- **Policies for Data Sharing: Numerous legal, ethical, security, and cultural constraints must be addressed to enable the requisite collection, aggregation, integration, analysis, interpretation and transmission of data for collaborative research. 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® Infrastructure Provides all such Requisites

caBIG® Provides Such a Translatable Informatics Platform to Support Individual and Institutional Needs

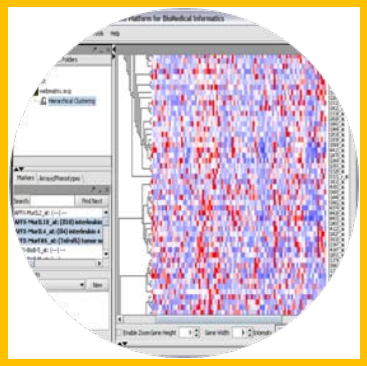


caBIG® Provides a Translatable Informatics Platform to Support Individual and Institutional Needs



caBIG® Provides a Translatable Informatics Platform to Support Individual and Institutional Needs

Analytical Tools

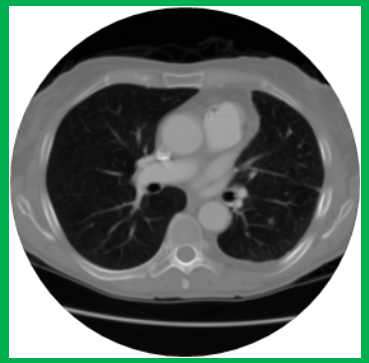


Molecular Analysis

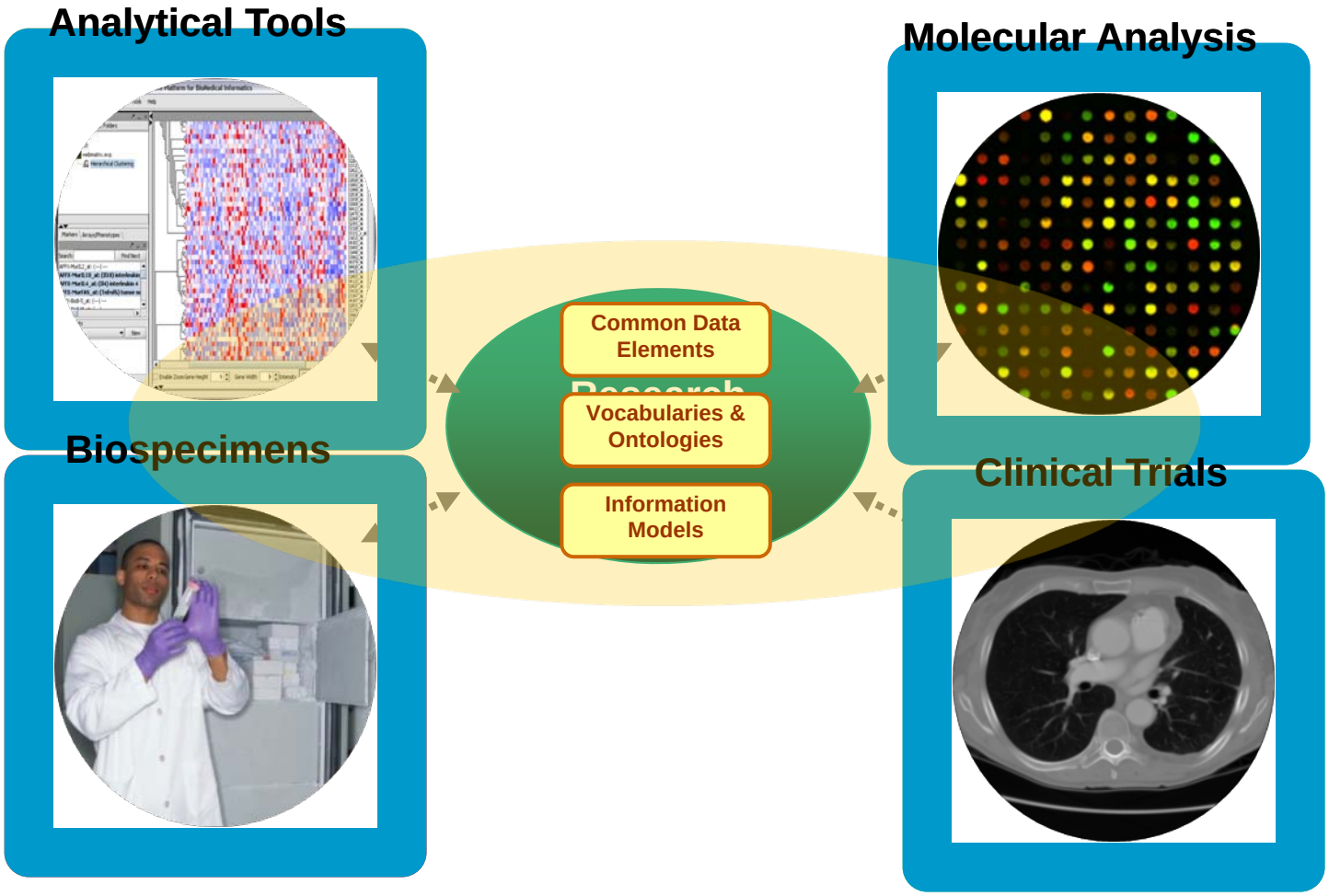
Research Unit

Biospecimens

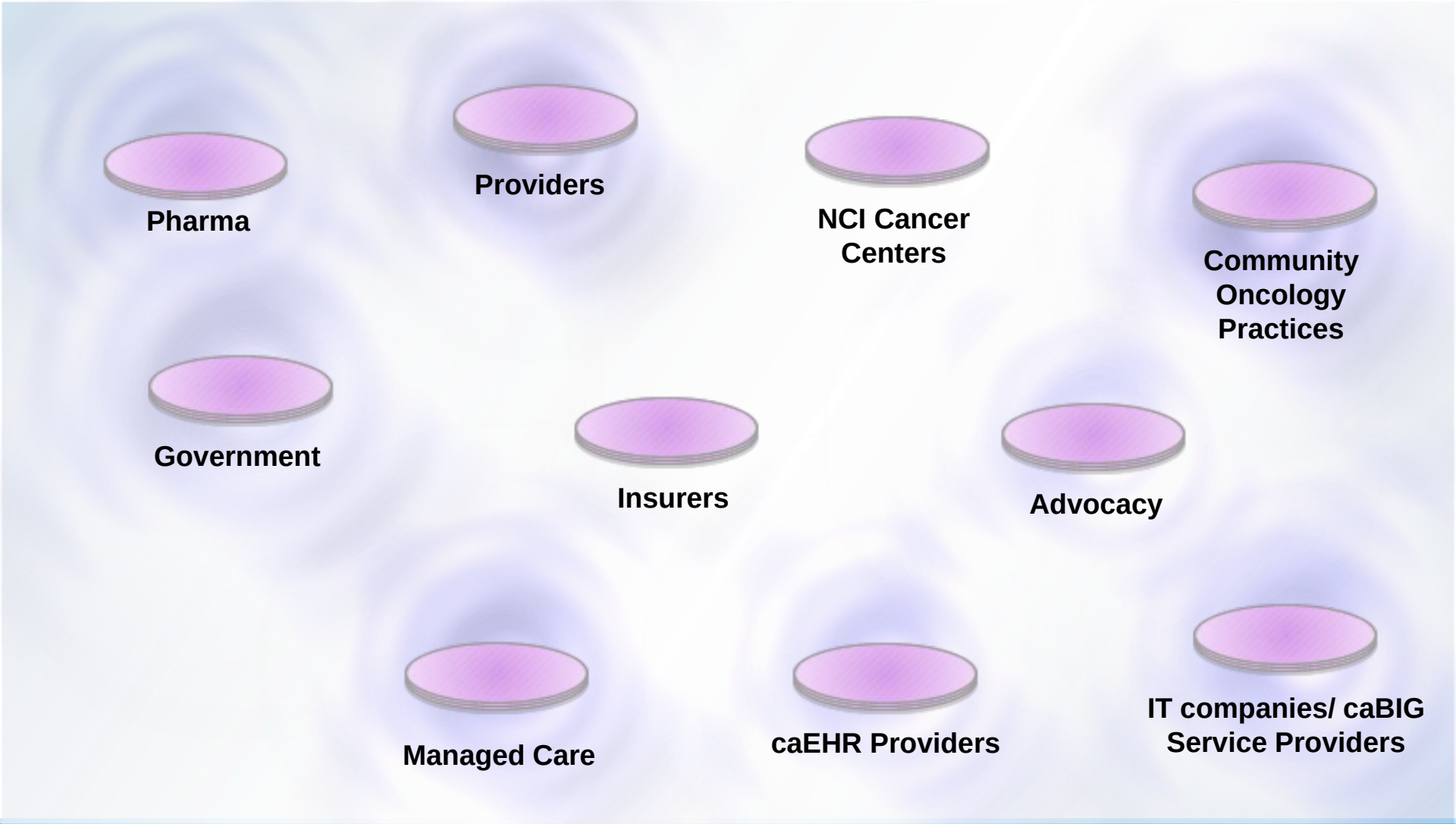
Clinical Trials



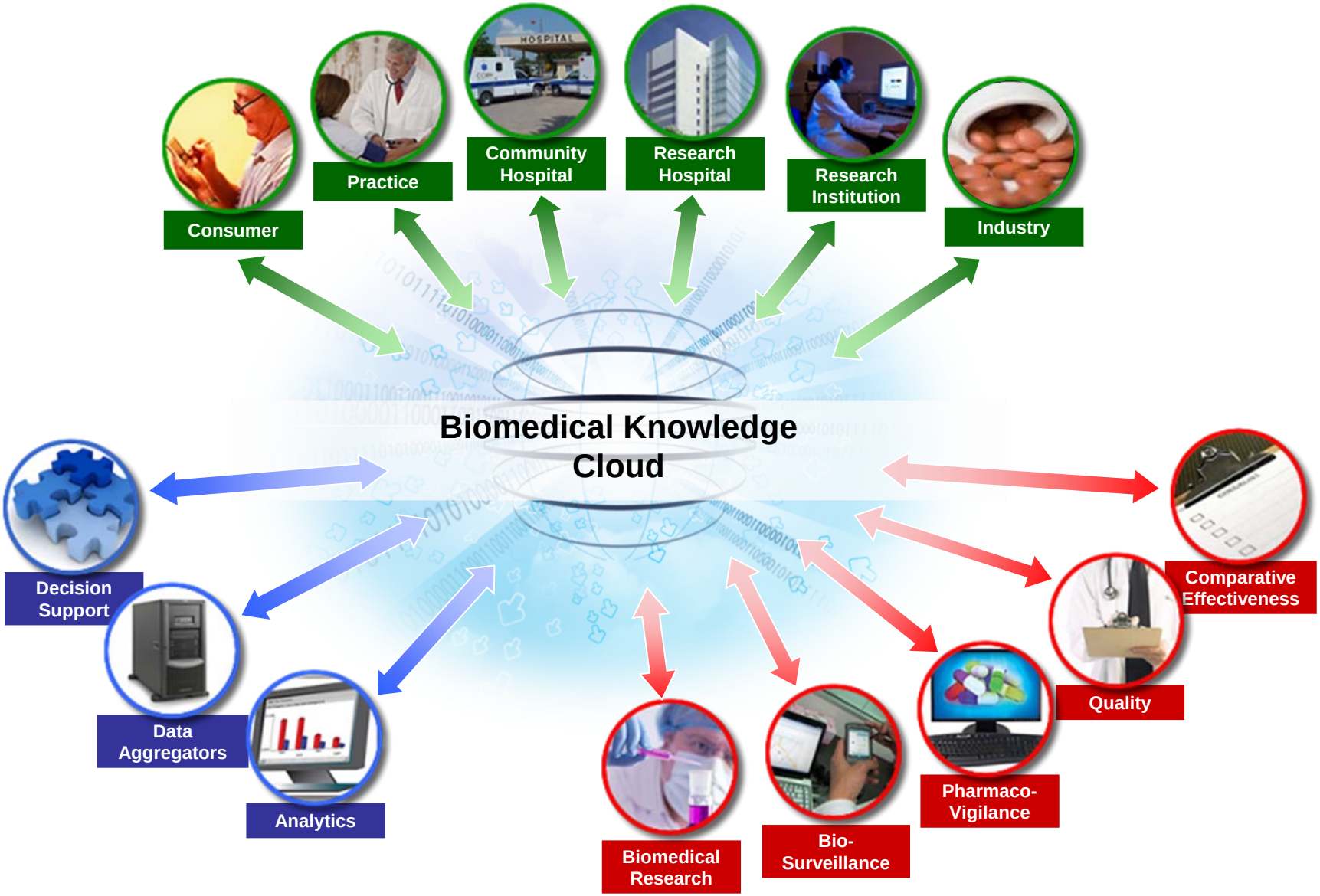
caBIG® Provides a Translatable Informatics Platform to Support Individual and Institutional Needs



Potential Participating Sectors in the Ecosystem



A 21st Century Biomedical Ecosystem



Potential Outcome of a Biomedical Ecosystem Enabled by Translatable Informatics

- **Faster discovery**
- **More productive product development**
- **Improved patient care**
- **Reduced healthcare costs**
- **Improved commercial opportunities**
- **Productive leveraging of a government investment**
- **Reduced government expenditures going forward**