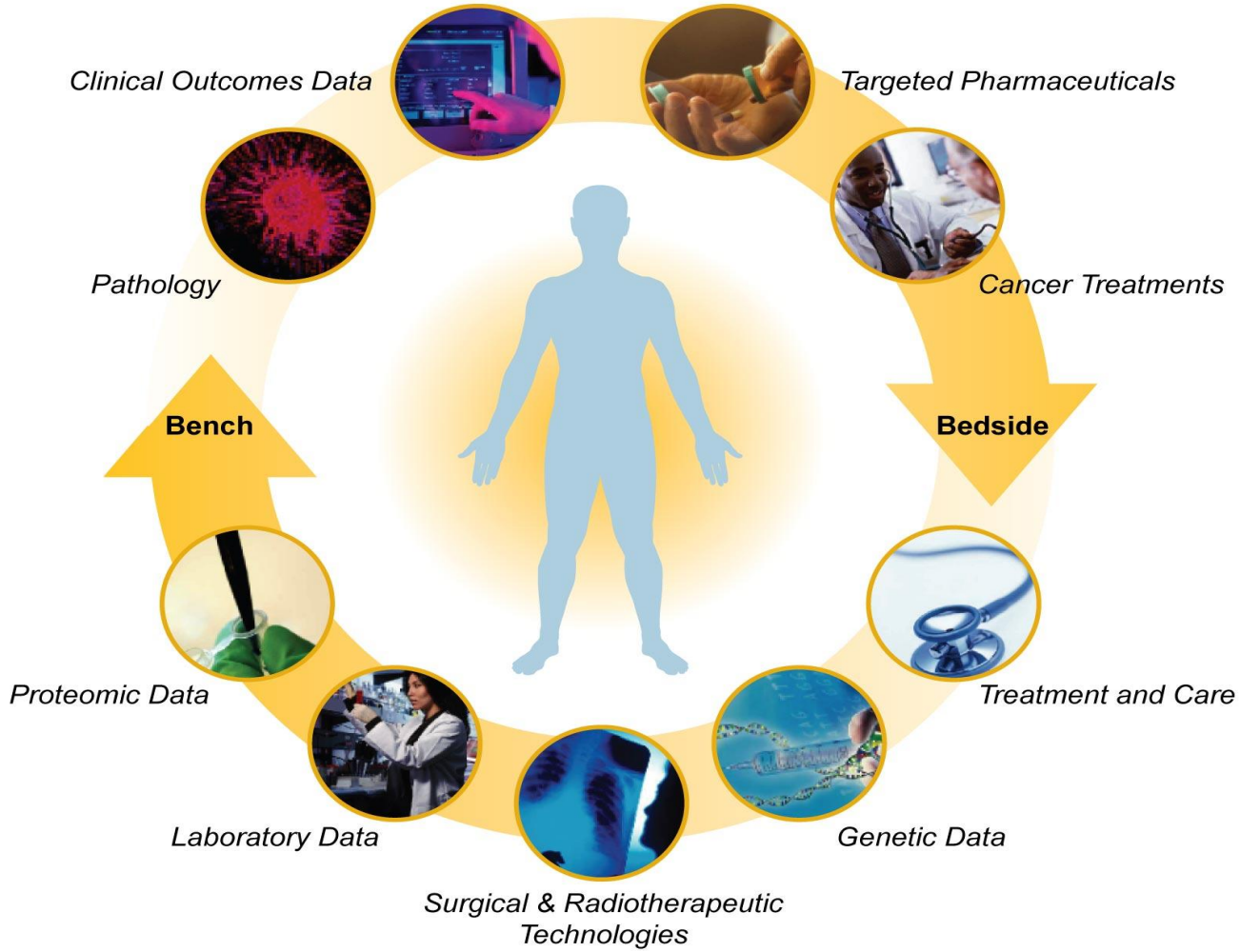
caBIG®
cancer Biomedical Informatics Grid®

Building a 21st Century Biomedical System: The Cancer Biomedical Informatics Grid (caBIG®)

Ken Buetow
National Cancer Institute

21st Century Medicine: What We're Trying to Achieve



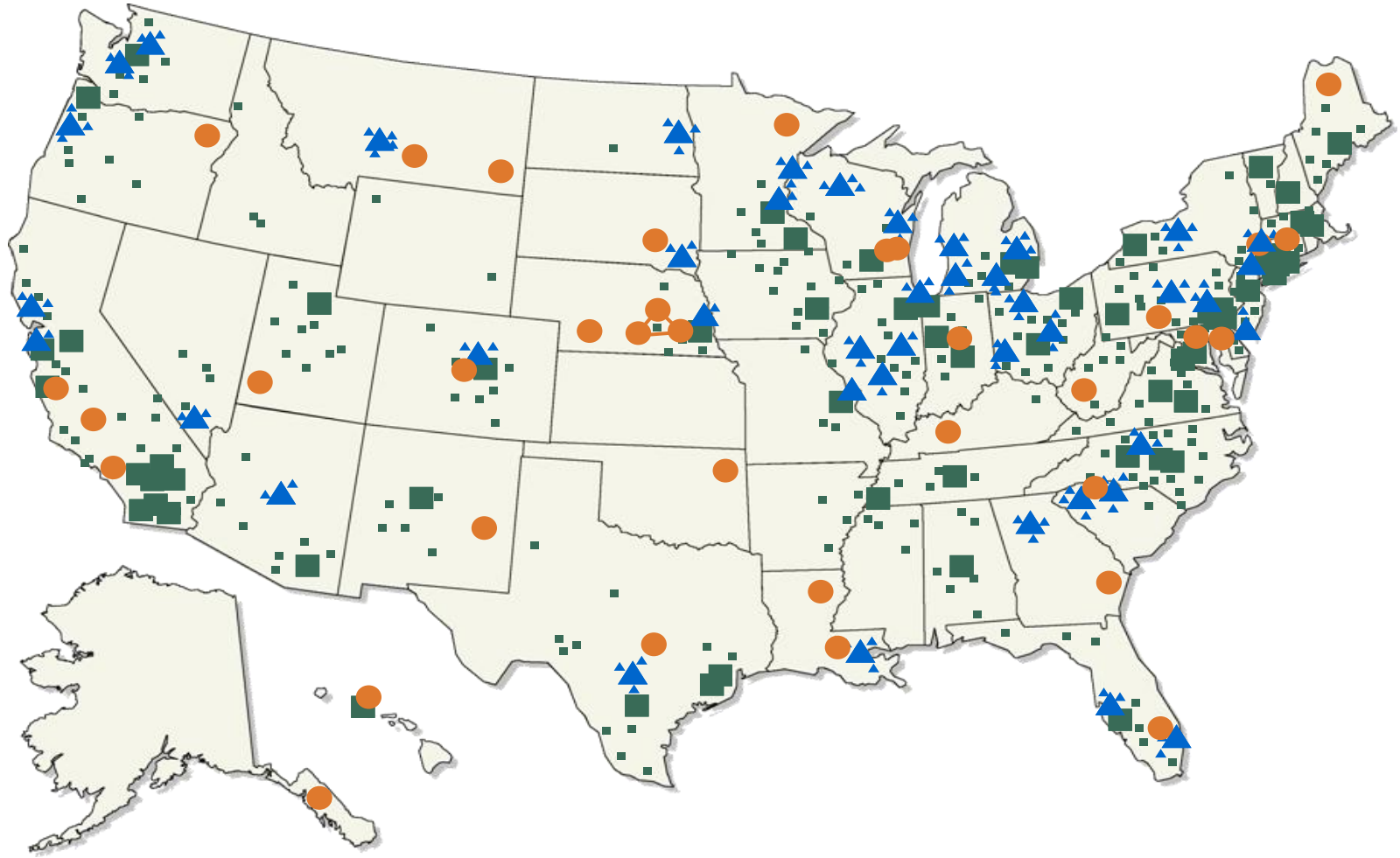
- **Personalized, Predictive, Preemptive, Participatory……**
- **Unifies clinical research, clinical care, and discovery (bench-bedside-bed) into a seamless continuum**
- **Results in improved clinical outcomes**
- **Accelerates the time from discovery to patient benefit**
- **Enables a health care system, not a disparate "sector"**
- **Empowers consumers in managing their health over a lifetime**

caBIG

Molecular Medicine as a Complex Continuum

Molecular Medicine

NCI-Designated Cancer Centers, Community Cancer Centers, and Community Oncology Programs

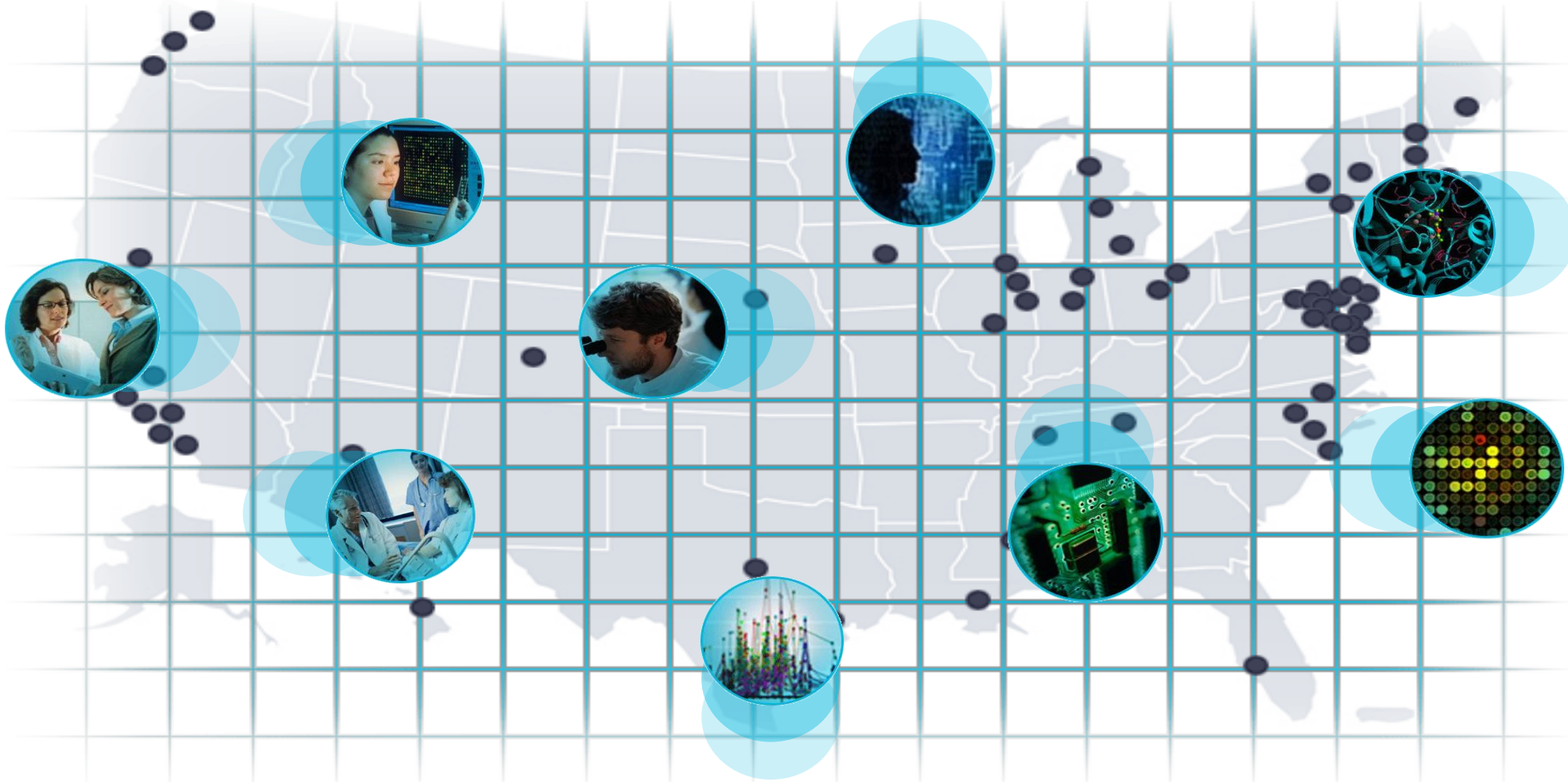


caBIG cancer Biomedical Informatics Grid™

Challenges:
The Biomedical Landscape

- **Isolated information “islands”**
- **Information dissemination uses models recognizable to Gutenberg**
- **Pioneered by Royal Academy of Science of London in the 17th century**
  - Write manuscripts
  - “Publish”
  - Exchange information at meetings

NCI is Utilizing Informatics to Integrate Cancer Information



caBIG cancer Biomedical Informatics Grid™

caBIG®
cancer Biomedical Informatics Grid®

Cancer Biomedical Informatics Grid (caBIG®): Providing the IT "Glue"

The caBIG® Initiative

caBIG® Goal

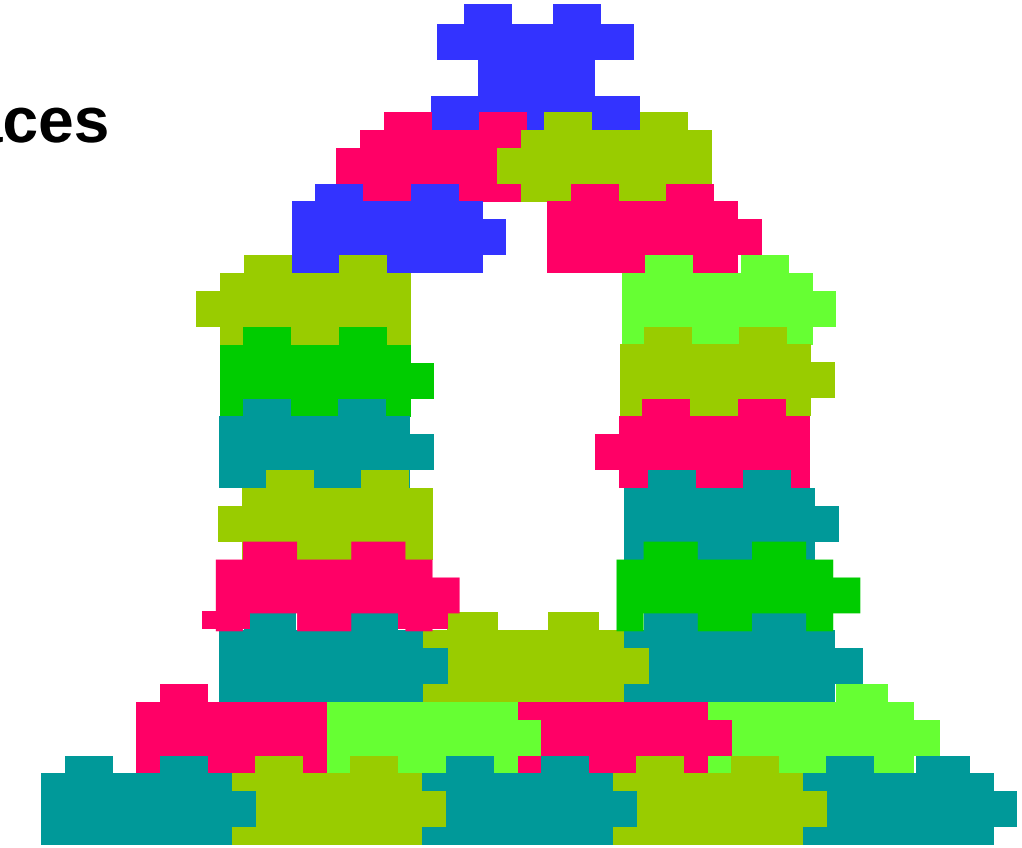
A virtual web of interconnected data, individuals, and organizations that redefines how research is conducted, care is provided, and patients/participants interact with the biomedical enterprise.

caBIG® Vision

- **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

caBIG® Approach

- **Modules that address specific needs**
- **Connect through defined Electronic interfaces**
- **Use of international data standards**



caBIG® Capabilities Enable

Discovery > Clinical Research > Clinical Care

Molecular Medicine

Clinical Research

Imaging

Molecular Biology

Pathology

caBIG® Capabilities Enable
Discovery > Clinical Research > Clinical Care

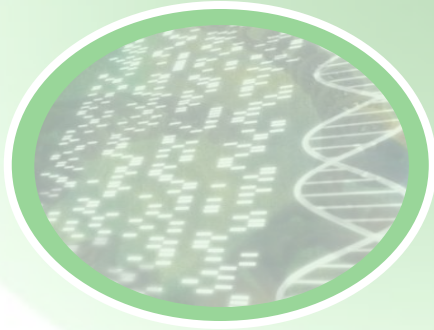
- Track clinical trial registrations
- Facilitate automatic capture of clinical laboratory data
- Manage reports describing adverse events during clinical trials

- 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 Medicine

Clinical Research

Imaging



Molecular Biology

Pathology

- 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

- 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

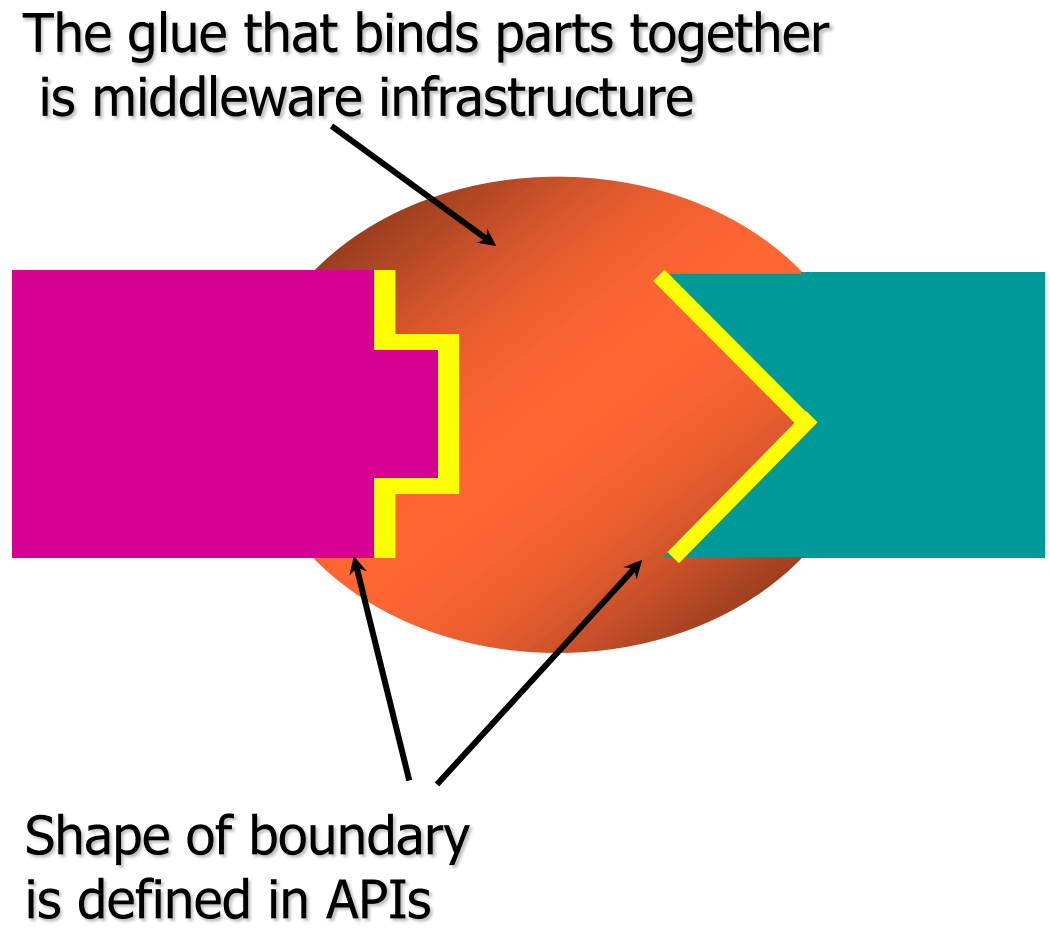
caBIG

caBIG®
cancer Biomedical Informatics Grid®

caBIG®: An open SOA with shared community semantics

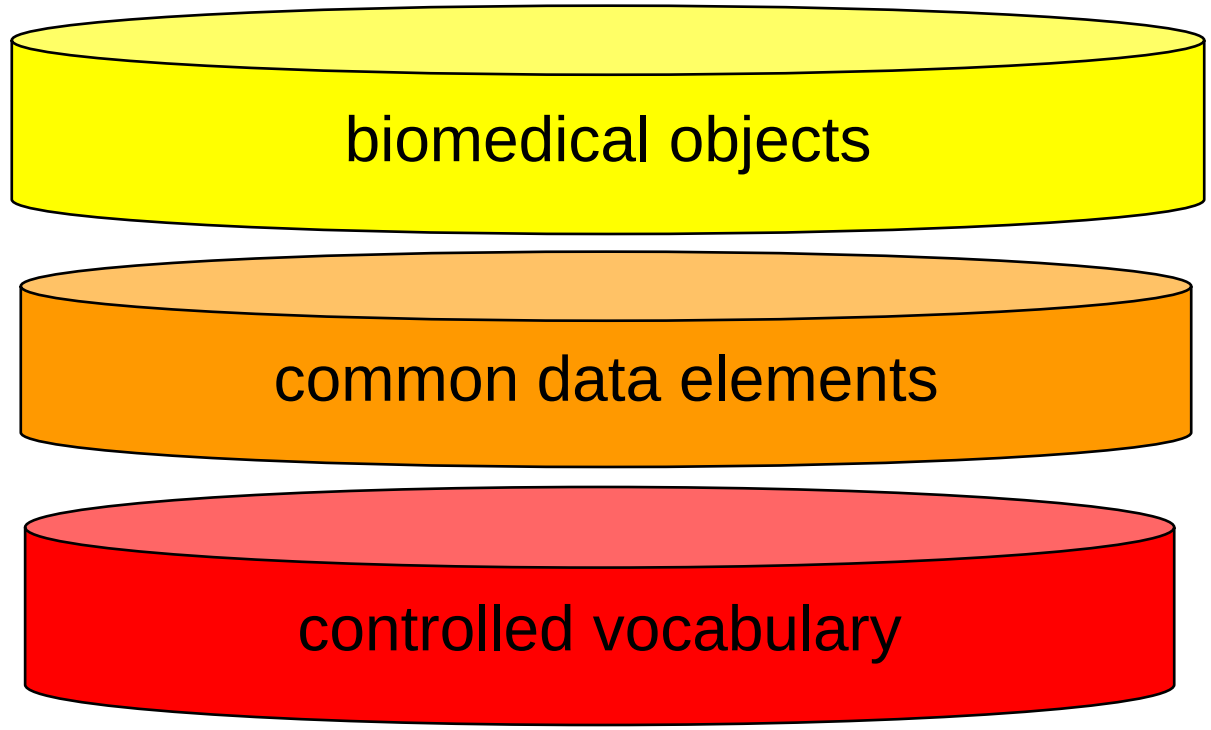
Boundaries and Interfaces

- **Focus on boundaries and interfaces, how things fit together, not on the internal details**
- **Once they're built: assume that will be diverse & changing**



Standards-Based Interoperability: caCORE

- **Community driven**
- **Dynamic implementation**
  - Built to be upgraded as standards “harden”, and domains expand

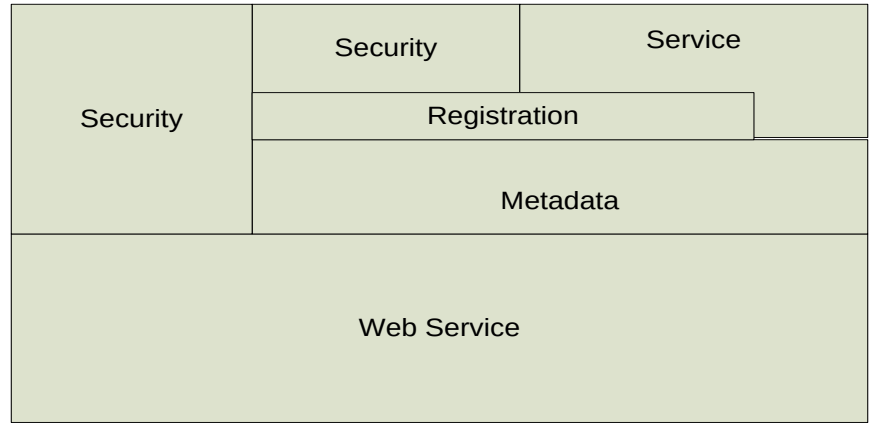


Standards Based Technical Infrastructure

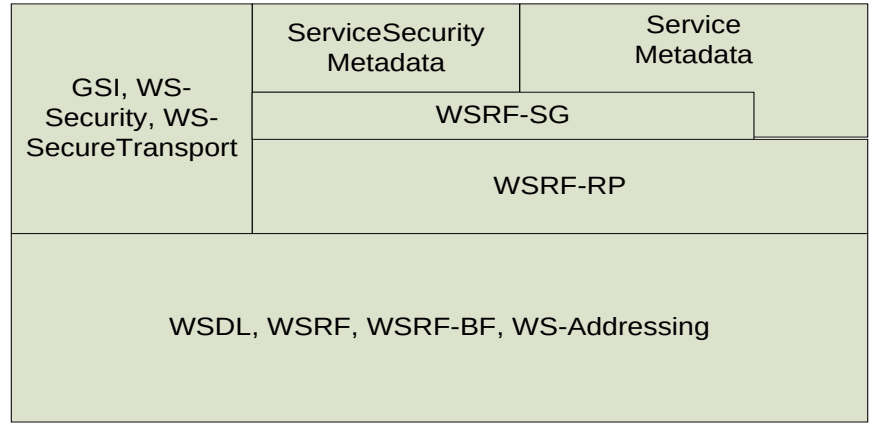
Grid Services

- caGrid uses the Globus Toolkit and Axis for creating, registering, discovering, and invoking these service operations as grid services
- Client uses the operation through a grid service interface and does not need to be aware of any implementation specific details of the grid service

Service Layers



Layer Specifications



Service Layers

Web Server (Apache/Tomcat): Binds to server port(s)

Web Application Server (Tomcat): Hosts web applications connected to the web server

SOAP Engine (Axis): Interprets SOAP requests, installed as a web application

Web/Grid Service (Globus): Binds "protocol" to operations on local application resources

Security (GSI)
* Secure Communication
* Authentication
* Authorization

Metadata (WSRF – Resource Properties)
* caGrid Service Metadata
* caGrid Service Security Metadata
* (caGrid Data Service Metadata)
* (Custom Metadata)

Service Definitions
* WSDL
* XSDs

Service Implementation

Advertisement (WSRF-SG)

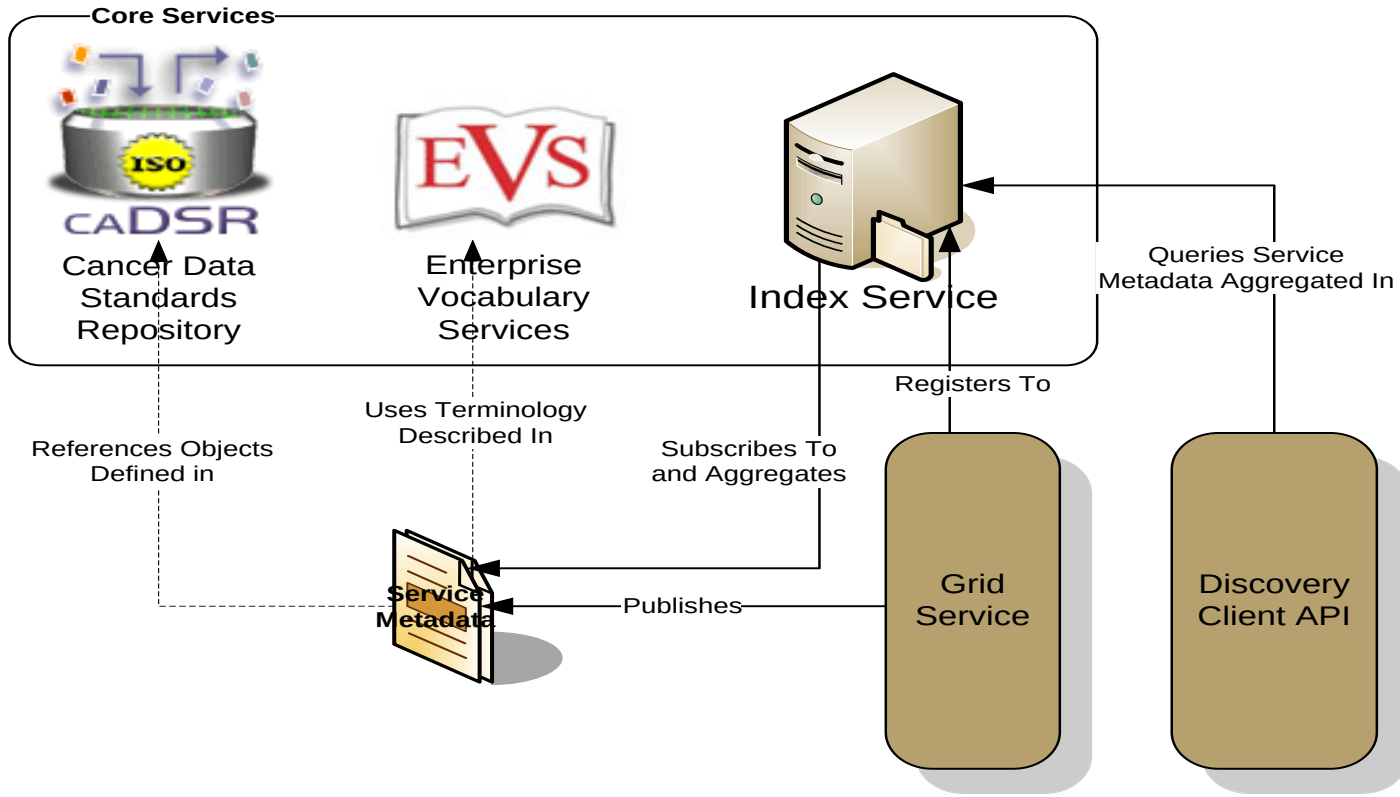
Configuration Properties

Business Logic

Resources (WSRF Resource)

Advertisement and Discovery Process

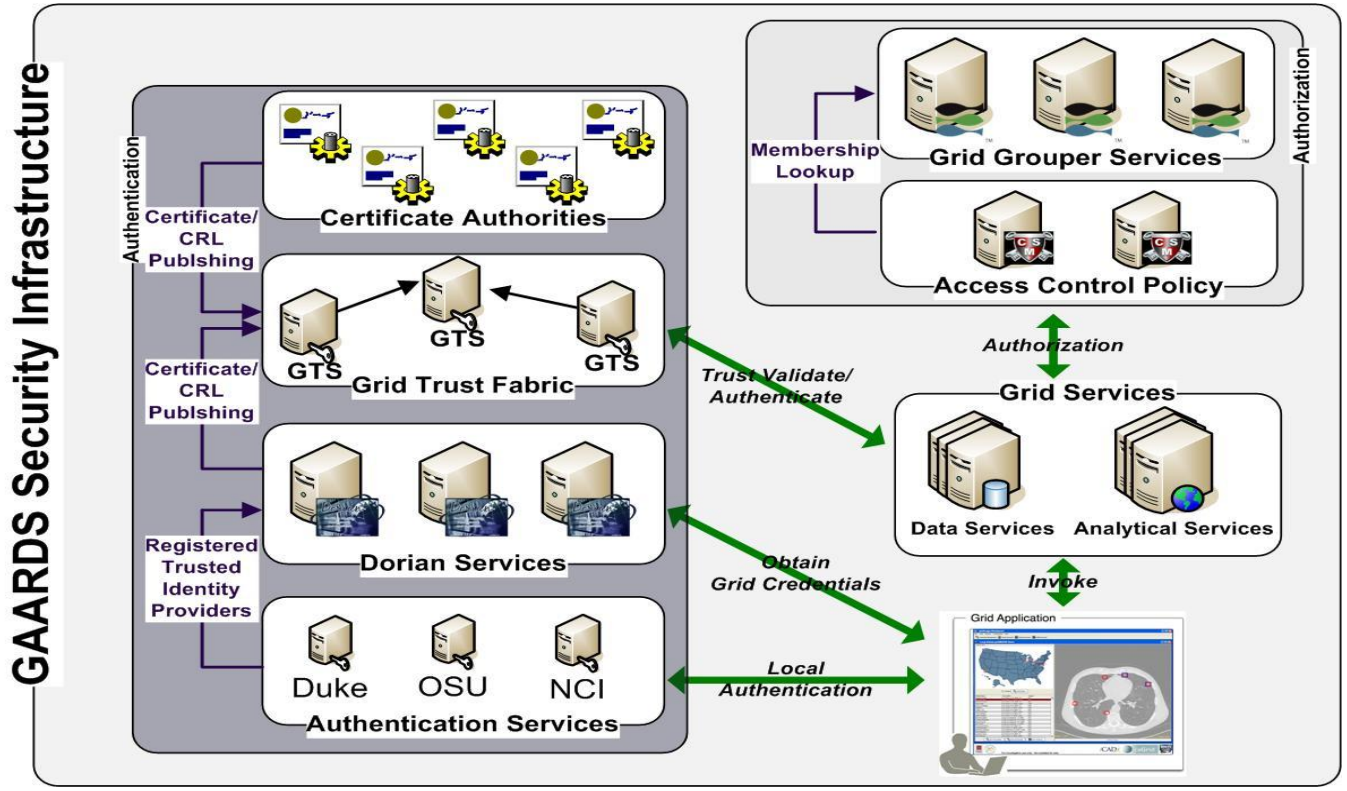
- All services register their service location and metadata information to an Index Service
- The Index Service subscribes to the standardized metadata and aggregates their contents
- Clients can discover services using a discovery API which facilitates inspection of data types
- Leveraging semantic information in EVS (from which service metadata is drawn), services can be discovered by the semantics of their data types



caGrid Security Infrastructure (GAARDS)

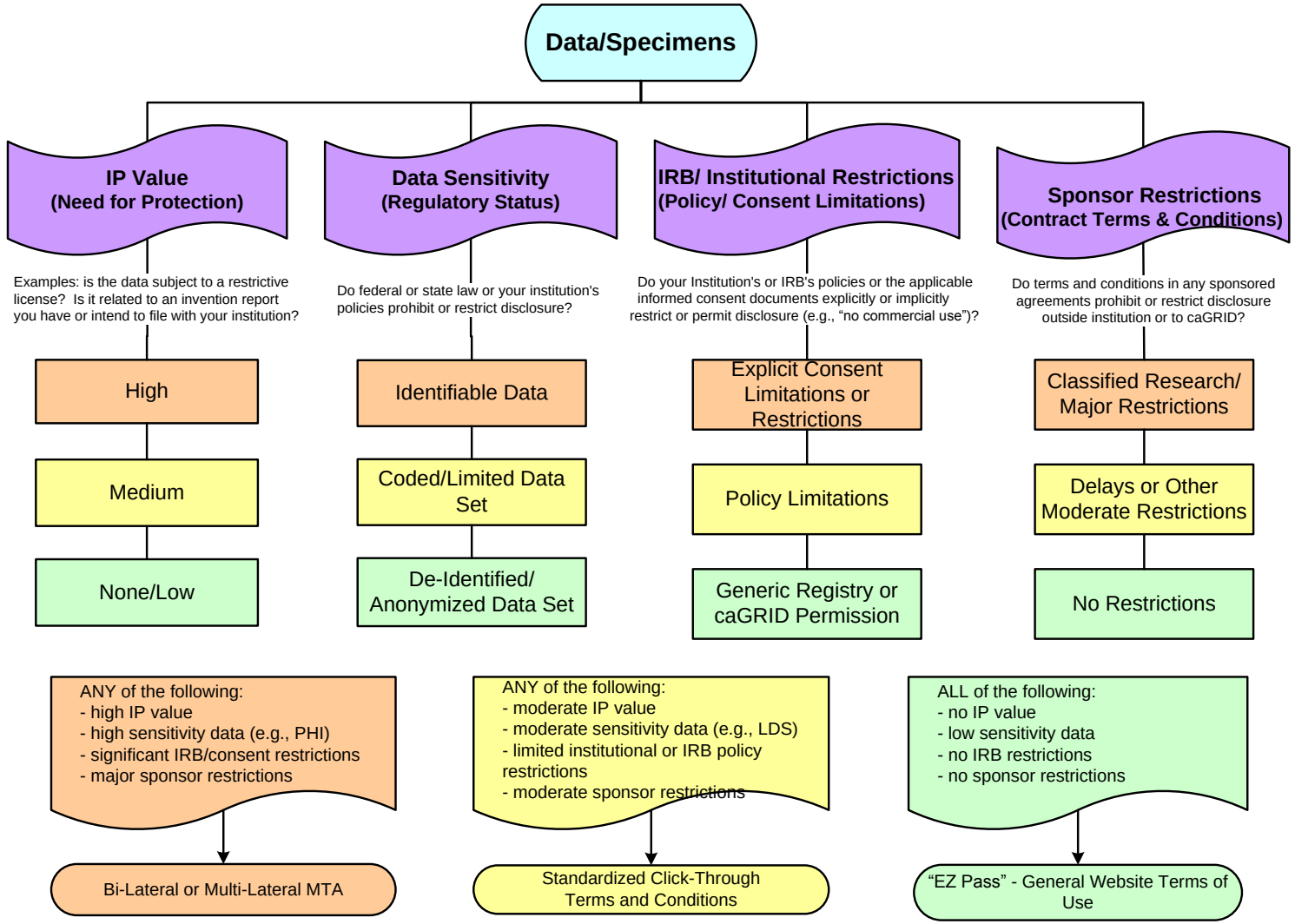
GAARDS provides services and tools for the administration and enforcement of security policy in an enterprise Grid.

- **Dorian**
  - Grid Account Management
  - Allows accounts managed in external domains to be federated and managed in the grid.
  - Dorian allows users to use their existing credentials (external to the grid) to authenticate to the grid
  - Issues Host Certificates.
- **Grid Trust Service (GTS)**
  - Creation and Management of a federated trust fabric.
  - Supports applications and services in deciding whether or not signers of digital credentials can be trusted.
  - Supports the provisioning of trusted certificate authorities and corresponding CRLS.

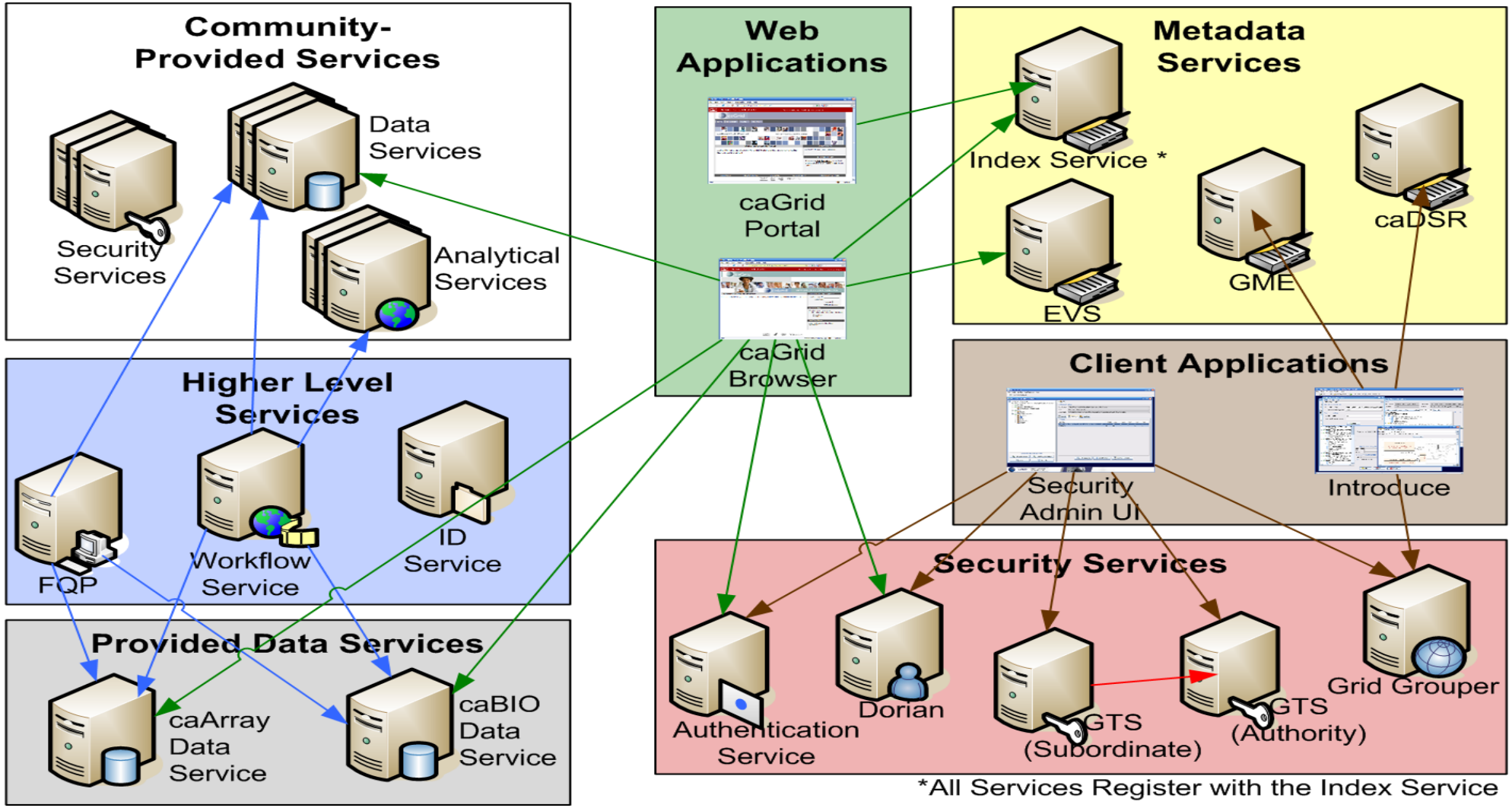


Policy: Analysis Framework

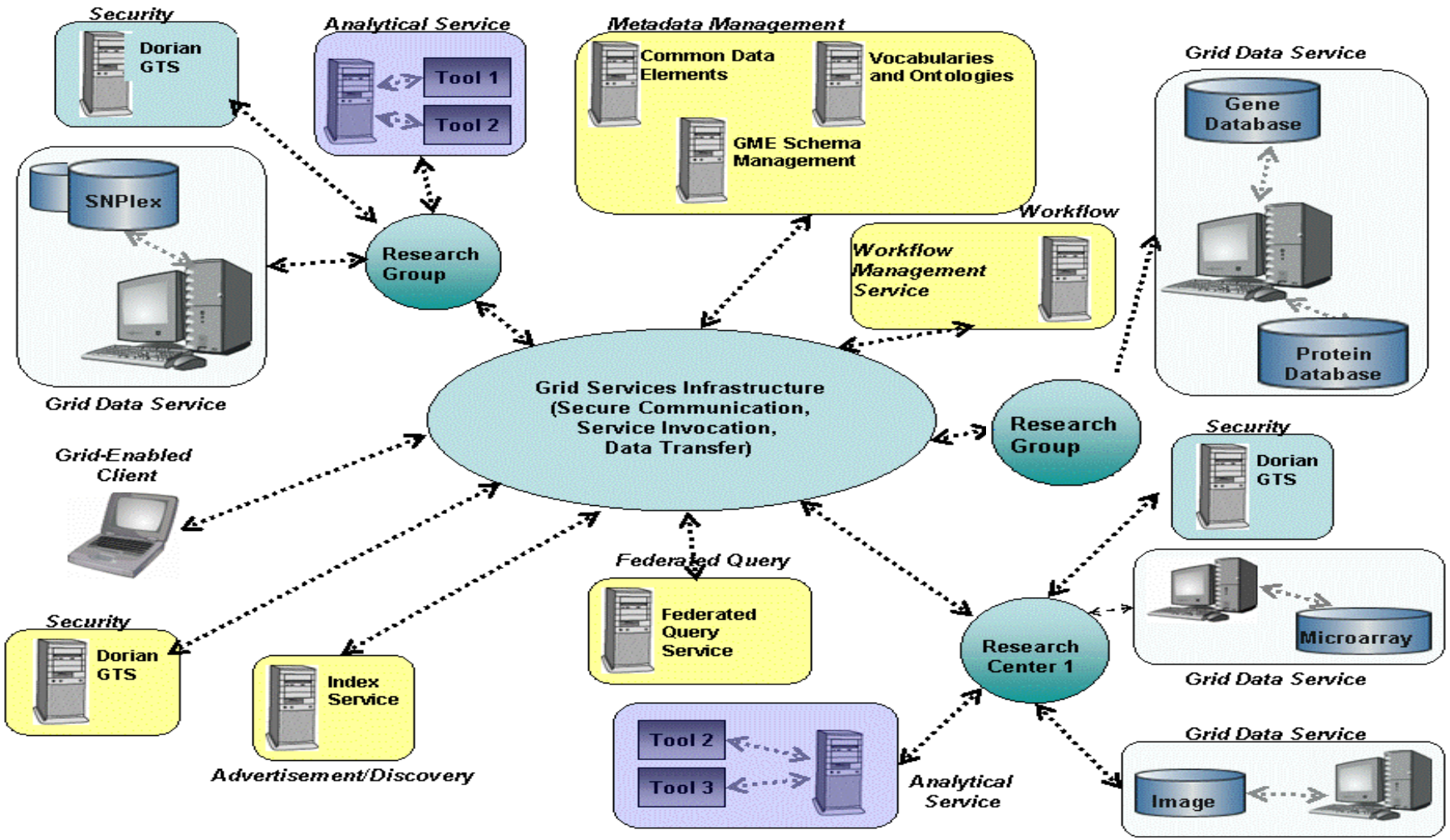
Decision Tree for Privacy/Intellectual Capital Terms and Conditions



caGrid Production Environment



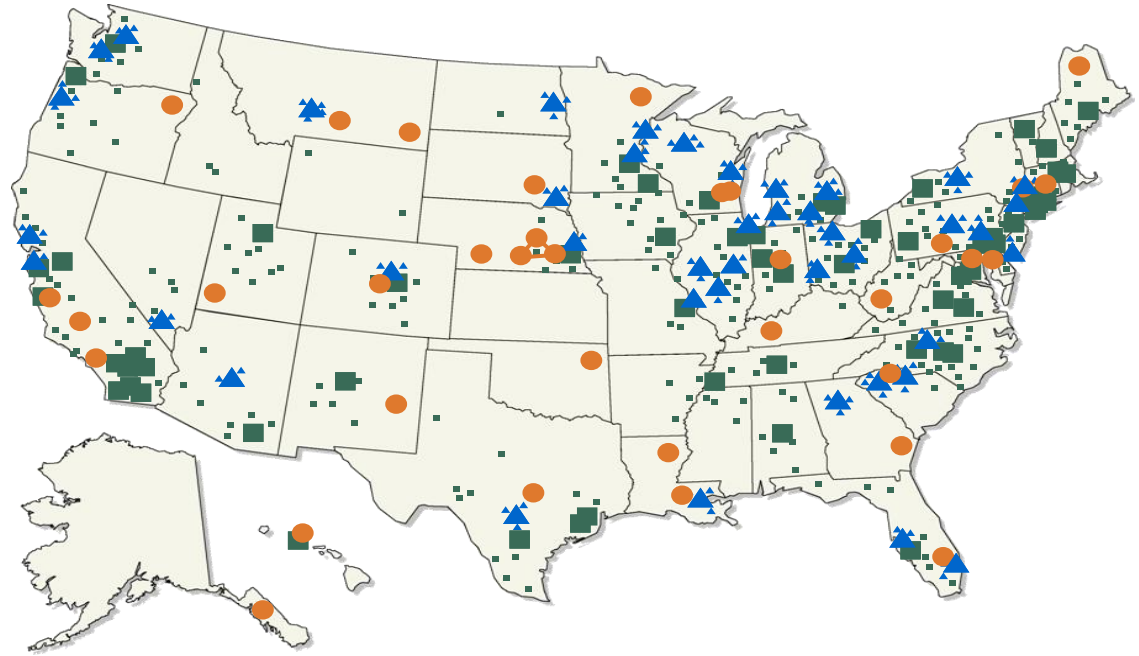
caGrid Conceptual View



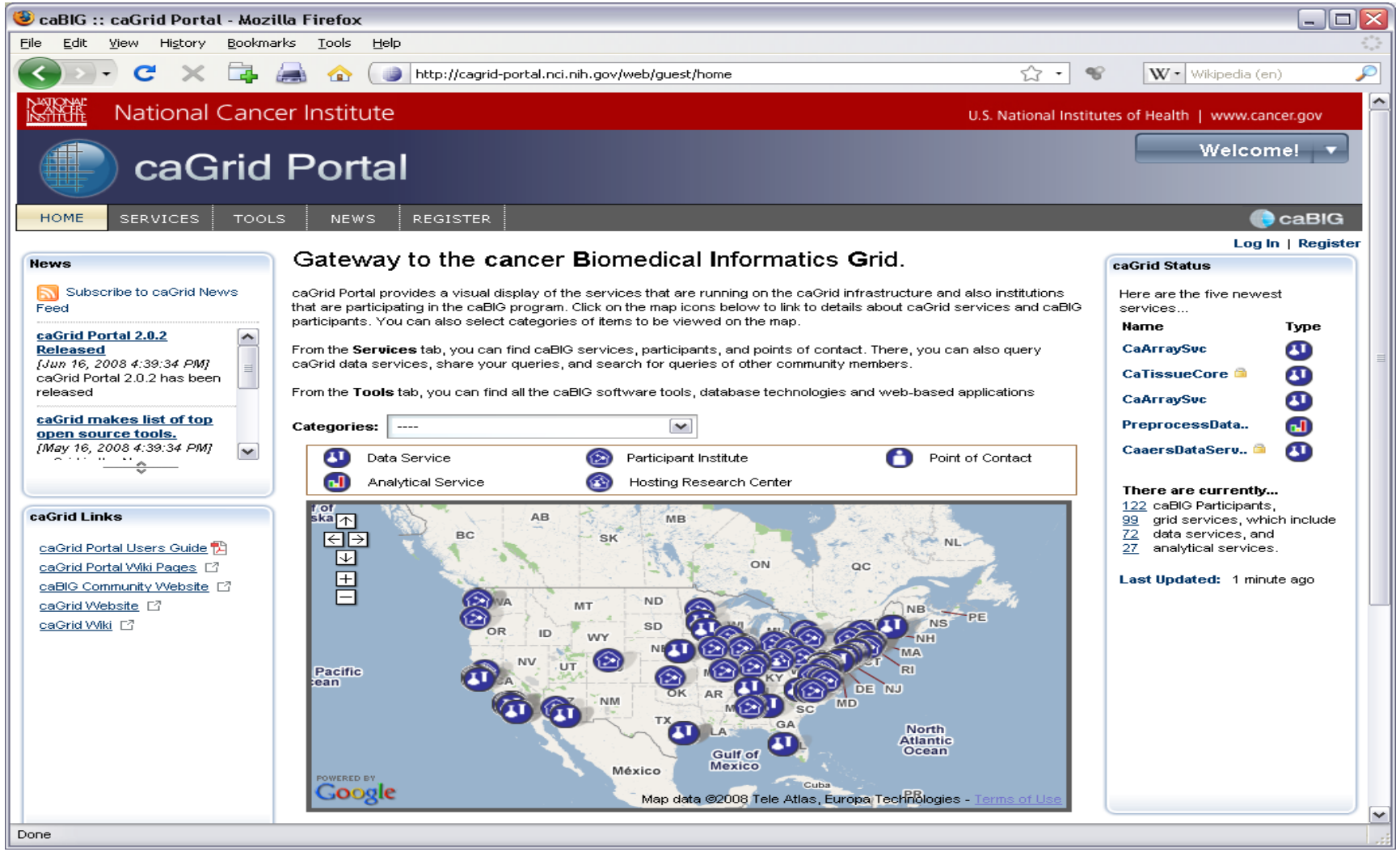
Connected with caBIG®

- **caBIG® adoption is unfolding in:**
  - 56 NCI-designated Cancer Centers
  - 16 NCI Community Cancer Centers
- caBIG® being integrated into federal health architecture to connect **National Health Information Network**
- **Global Expansion**
  - United Kingdom
  - China
  - India
  - Latin America

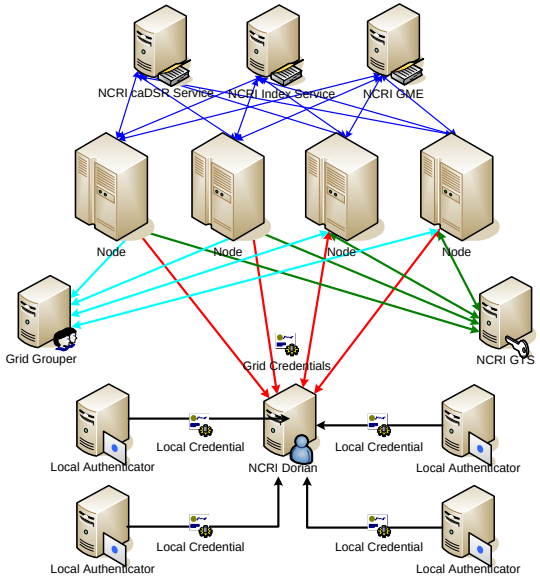
NCI-Designated Cancer Centers, Community Cancer Centers, and Community Oncology Programs



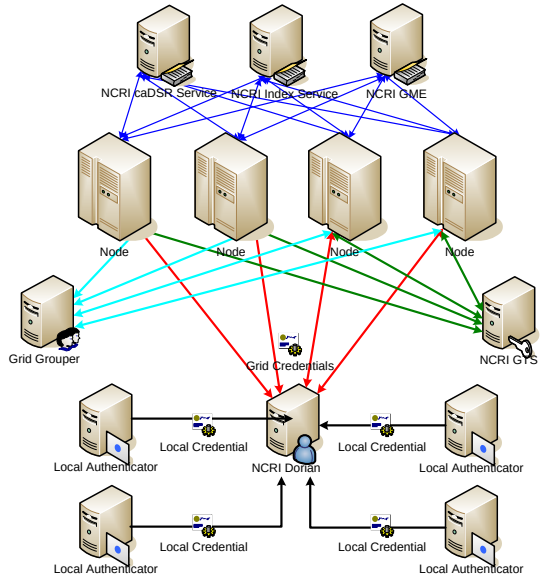
caBIG® Snapshot



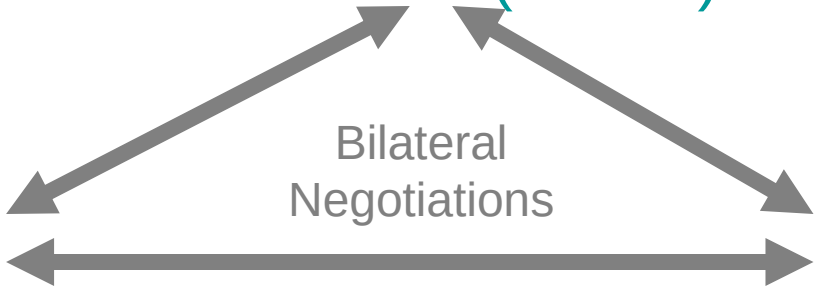
Grid of Grids…

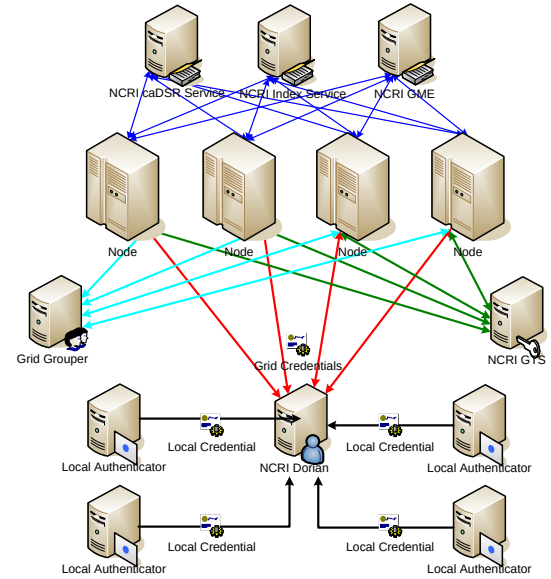


caGrid



NHLBI Grid (CVRG)





NCRI ONIX

caBIG

caBIG® - The Next Wave:

Reaching End Users

caBIG® Installation Support

- A software packaged to facilitate installation including on-lineTutorials and Videos and Learning Center materials

Knowledge Centers – Power to the People

- A one-stop shop for domain expertise, technical and end-user documentation, training materials, and comprehensive, up-to-date installation packages for caBIG® tools
- A centralized site providing administration of open source development of caBIG® tools, collection and monitoring of defect reports, feature requests, and end-user requirements

Licensed Support Service Providers – Helping Hands

- Organizations that provide hands-on deployment support, customization, help desk, documentation development and end-user training

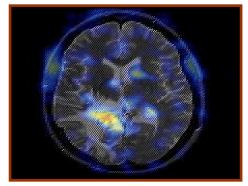
caBIG®
cancer Biomedical Informatics Grid®

Case Study: The Cancer Genome Atlas (TCGA)

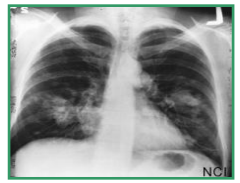
Connecting multiple sources, experiments, and data types

Three forms of cancer

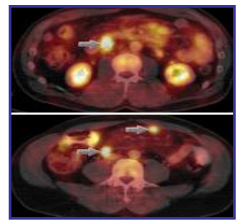
glioblastoma multiforme (brain)



squamous carcinoma (lung)



serous cystadenocarcinoma (ovarian)

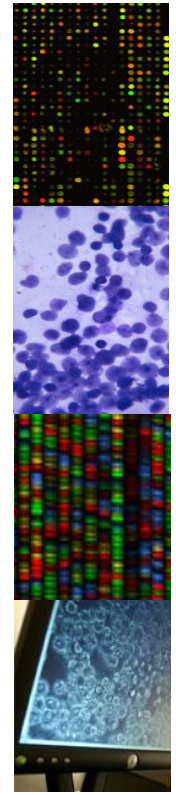


12 Organizations

- Biospecimen Core Resource
- 7 Cancer Genomic Characterization Centers
- 3 Genome Sequencing Centers
- Data Coordinating Center

Multiple data types

- Clinical diagnosis
- Treatment history
- Histologic diagnosis
- Pathologic status
- Tissue anatomic site
- Surgical history
- Gene expression
- Chromosomal copy number
- Loss of heterozygosity
- Methylation patterns
- miRNA expression
- DNA sequence



Empowering researchers to integrate increasingly complex layers of cancer biology, from gene to clinical phenotype, as a whole

Gene ⟷ Genome ⟷ Genomes ⟷ Pathways ⟷ Clinical Outcomes

All from their computer

A single web-based portal for all analyses **– http://cma.nci.nih.gov**

National Cancer Institute

Context: TCGA

caBIG™ cancer Biomedical Informatics Grid ™

Cancer Molecular Analysis Portal

Gene View

Genome View

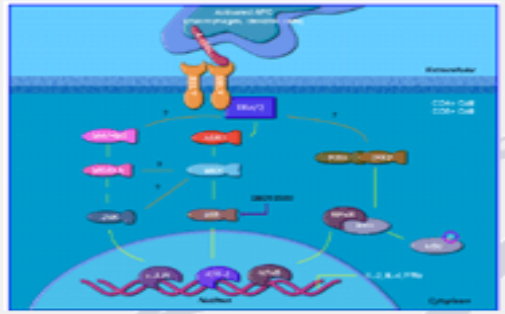
Clinical View

Analysis Tools

Gene View

Visualize gene expression, copy number, SNP, and pathway data on a gene by gene basis. Generate detailed study related reports for a given gene.

Available resources include: Gene Expression Plots, KM Survival Plots, CGWB Integration, and Pathway Visualizations.



Existing Users:

user:

pass:

login

Additional Information:

- Register
- Provide your feedback

NATIONAL CANCER INSTITUTE · FIRSTGOV

A single web-based portal for all analyses **– http://cma.nci.nih.gov**

National Cancer Institute

Context: TCGA

caBIG™ cancer Biomedical Informatics Grid ™

Cancer Molecular Analysis Portal

Gene View

Genome View

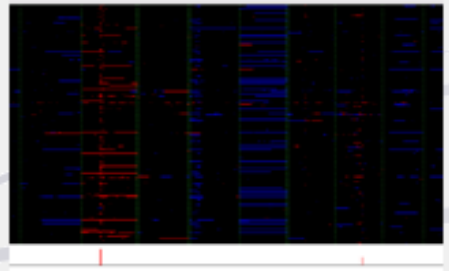
Clinical View

Analysis Tools

Genome View

Explore all of the study data in one genome level visualization. Investigate chromosomal regions of amplification, deletion and over expression. Zoom in on a chromosomal region of interest for a more detailed view.

Available resources include: Integrated Heatmap Viewer for Genomic Data.



Existing Users:

user:

pass:

login

Additional Information:

- **Register**
- **Provide your feedback**

FIRSTGOV

A single web-based portal for all analyses **– http://cma.nci.nih.gov**

National Cancer Institute

Context: TCGA

caBIG™ cancer Biomedical Informatics Grid ™

Cancer Molecular Analysis Portal

Gene View

Genome View

Clinical View

Analysis Tools

Clinical View

Investigate the study clinical data. Explore the relationships between clinical and molecular study data.

Available resources include: Clinical Query with Reports and KM Sample Plots

Existing Users:

user:

pass:

login

Additional Information:

- **Register**
- **Provide your feedback**

A single web-based portal for all analyses **– http://cma.nci.nih.gov**

National Cancer Institute

Context: TCGA

caBIG™ cancer Biomedical Informatics Grid ™

Cancer Molecular Analysis Portal

Gene View

Genome View

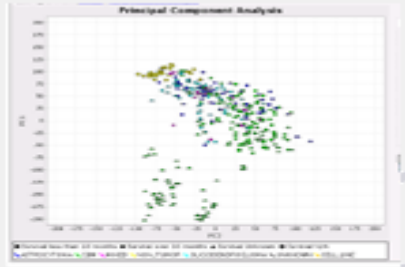
Clinical View

Analysis Tools

Analysis Tools

Analyze the study data using analysis tools such as Genepattern, Principal Component Analysis, and the Cancer Genome Workbench

Available resources include: PCA Analysis and Gene Pattern Integration



GenePattern

Existing Users:

user:

pass:

login

Additional Information:

- Register
- Provide your feedback

NATIONAL CANCER INSTITUTE FIRSTGOV

Comprehensive Summary

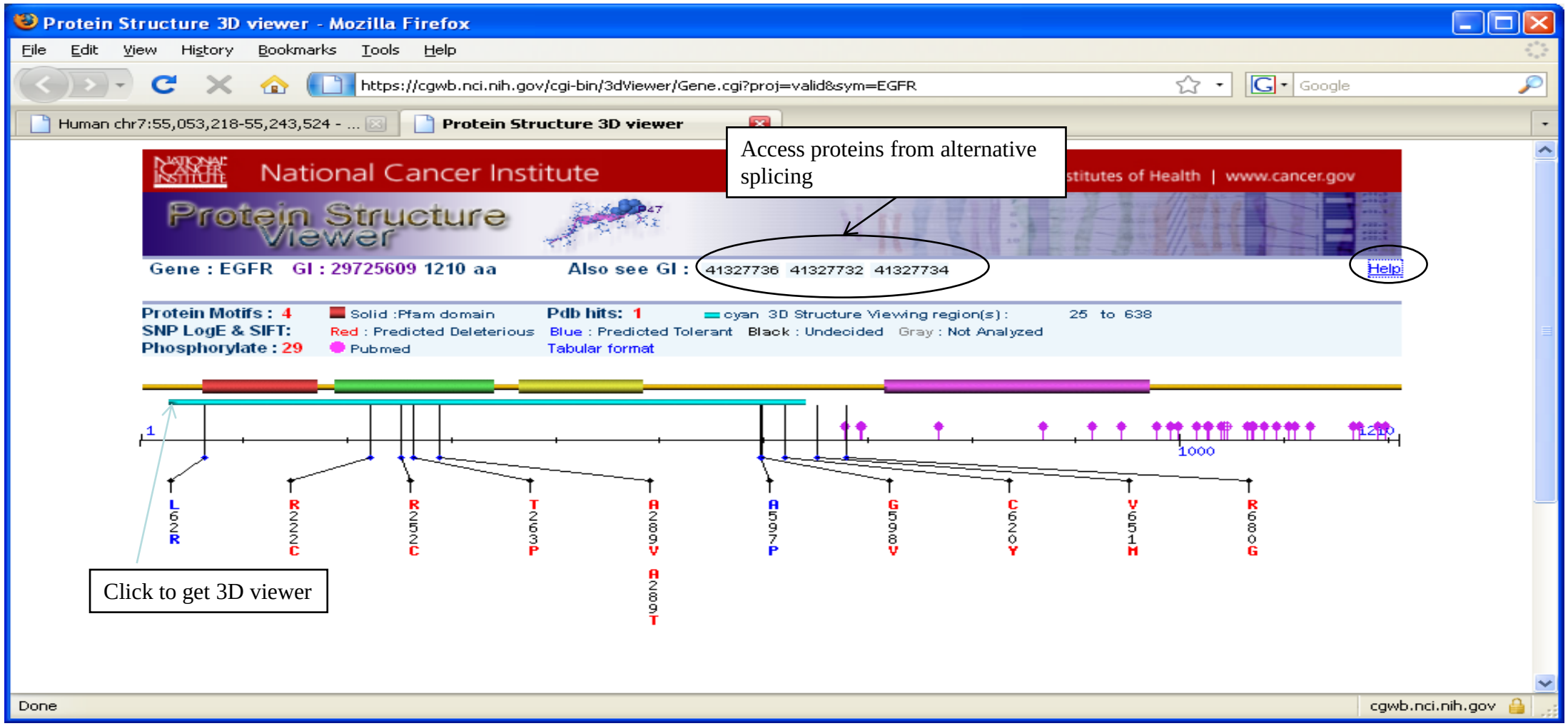
Putative Somatic Mutations Can Be Manually Reviewed

Example: A frameshift mutation in EGFR in paired tumor/normal

Normal

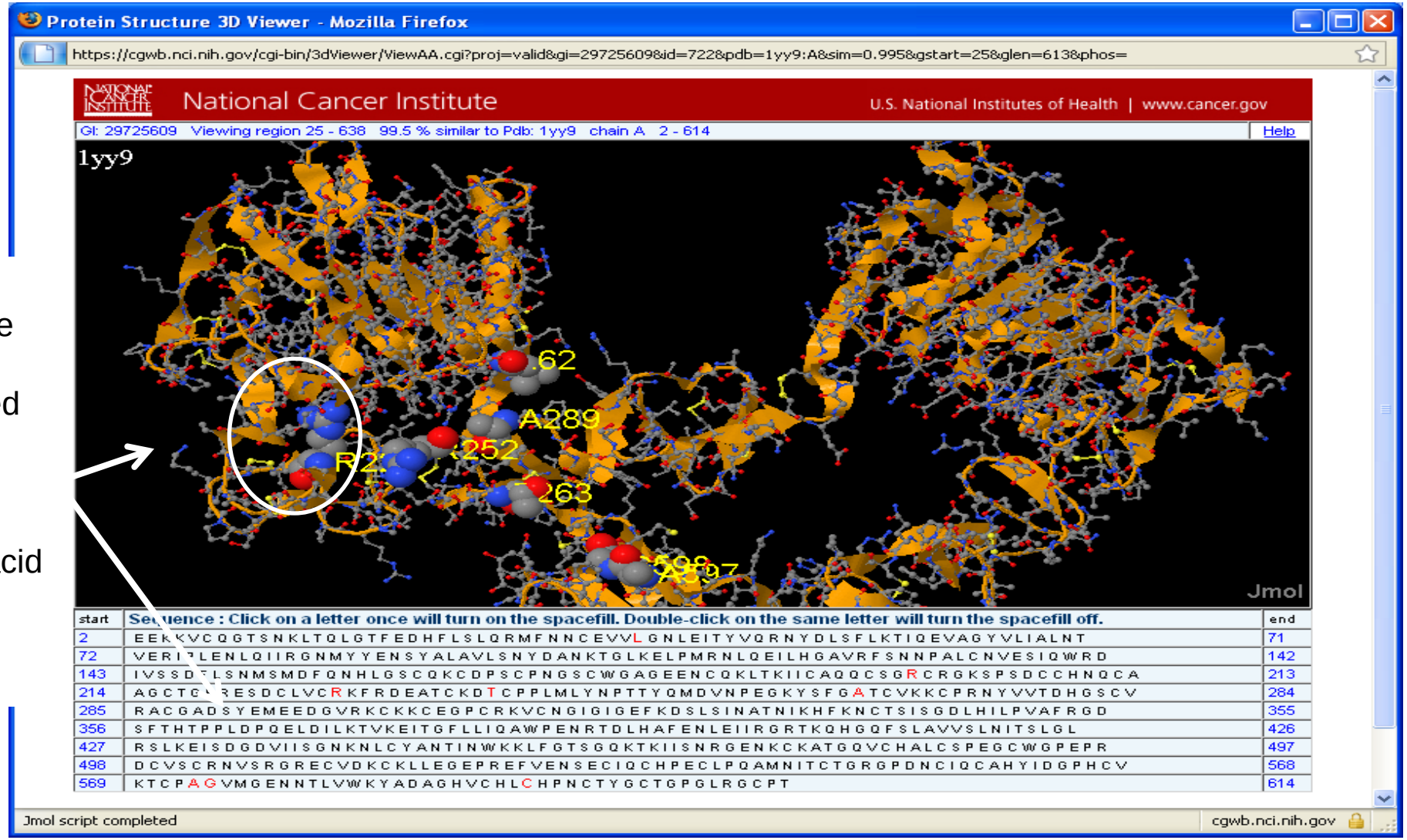
Tumor

Protein Structure View of EGFR Mutations

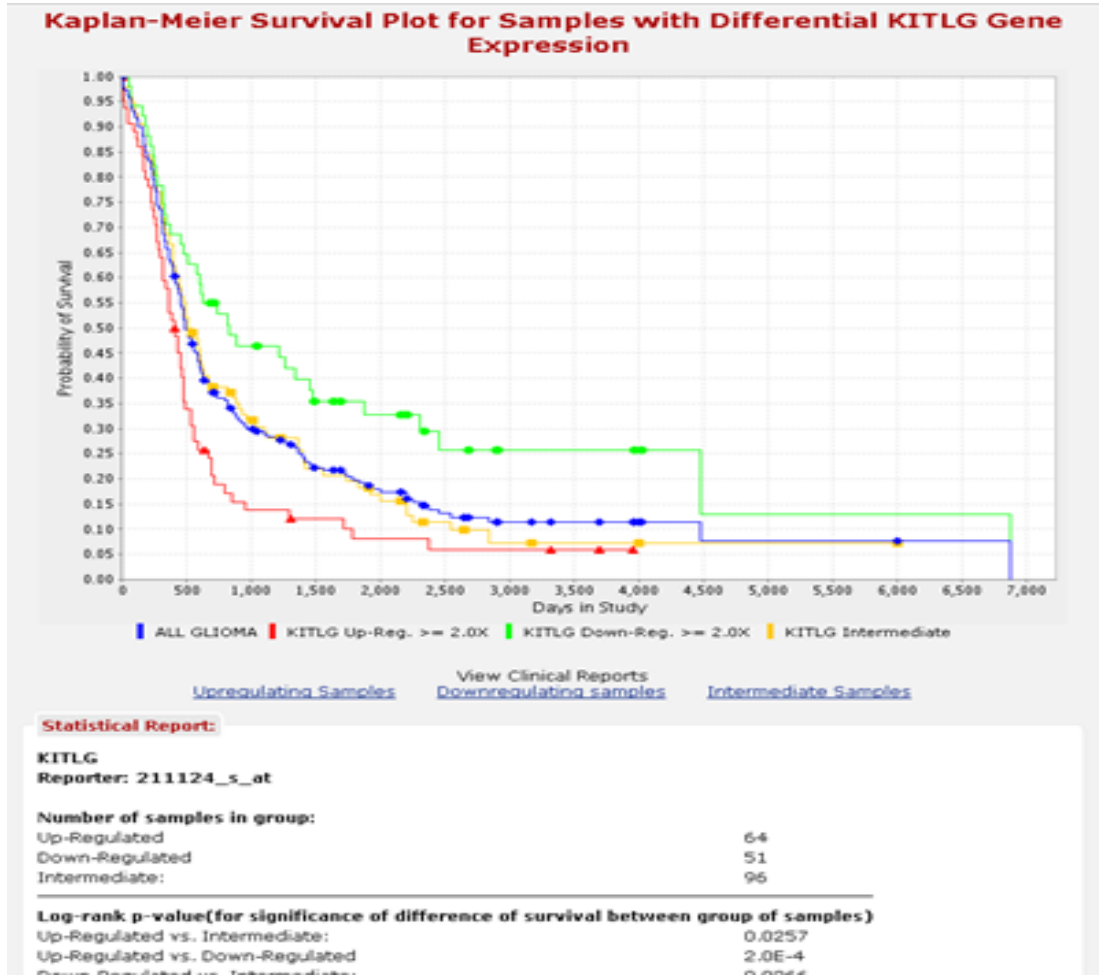


3D Structure Viewer

Big, highlighted atoms refer to the mutated amino acids (show in red in the bottom panel). You can also click on the mutated amino acid (shown in red) to turn on or off a specific mutation



Gene Expression Analyses Related to Clinical Outcome



View Clinical Reports
Upregulating Samples | Downregulating samples | Intermediate Samples

Statistical Report:

KITLG
Reporter: 211124_s_at

Number of samples in group:

Up-Regulated	64
Down-Regulated	51
Intermediate:	96

Log-rank p-value(for significance of difference of survival between group of samples)

Up-Regulated vs. Intermediate:	0.0257
Up-Regulated vs. Down-Regulated	2.0E-4

Administration:

- View Results
- List Management
- Help

News:

- Data Version
- TCGA newsletter - March 2008
- Number of Patients - 110
- Number of Expression Arrays - 985
- Number of Copy Number Arrays - 361

PatientDID Lists:

- ALL_PATIENTS
- Low_Survival
- Med_Survival
- High_Survival
- TP53_SomaticMu...
- EGFR_SomaticMu...
- PTEN_SomaticMu...
- RB1_SomaticMut...
- DST_SomaticMut...
- NF1_SomaticMut...
- CDKN2A_Somatic...
- PIK3R1_Somatic...
- CENPF_SomaticM...
- ITGB3_SomaticM...

Gene Lists:

- TCGA Target Se...

Reporter Lists:

High Order Analyses: Pathways

Pathways and Associated Anomalies

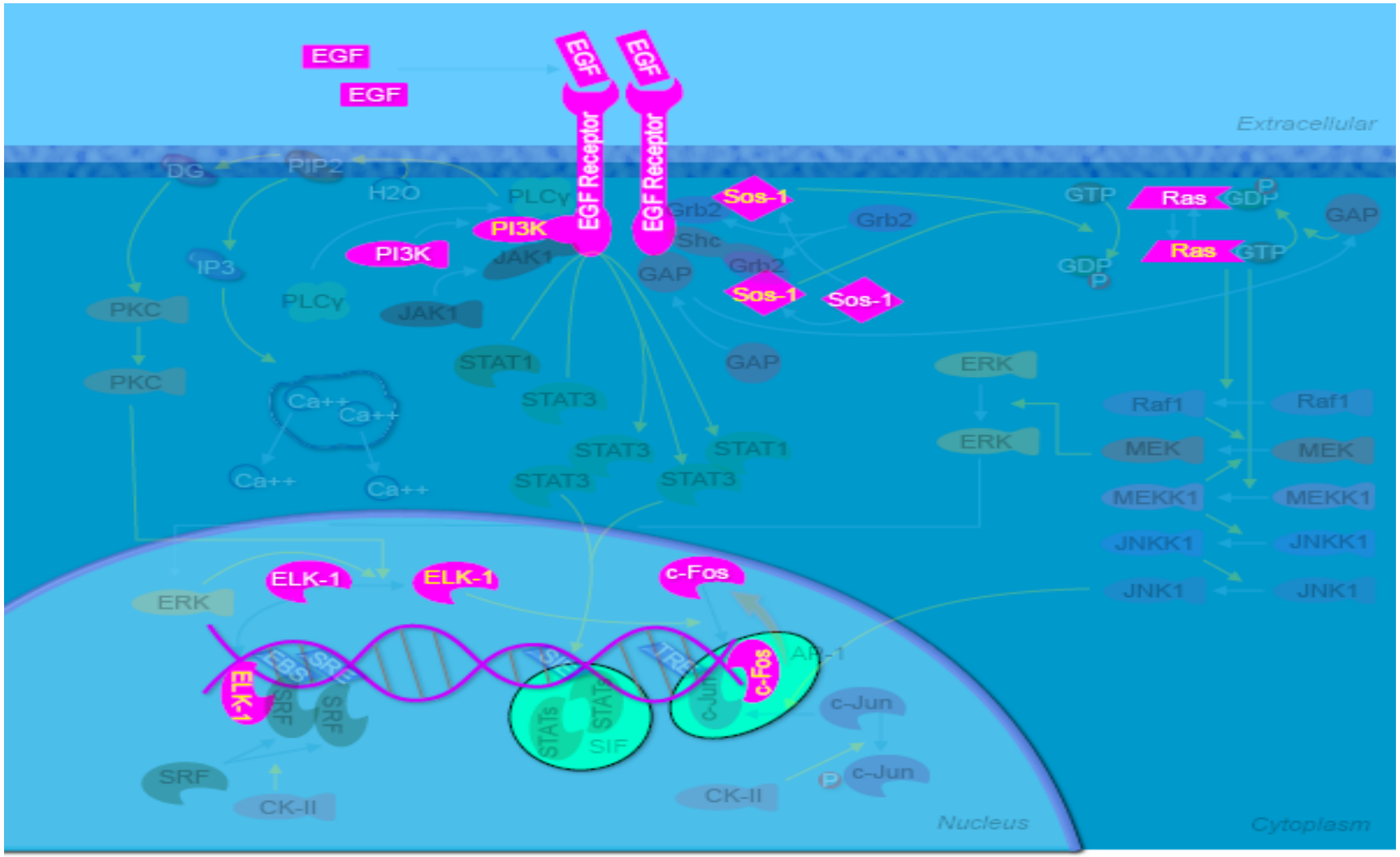
First Prev Next Last | 100 Rows Displayed

309 results found, displaying 1 to 100

Pathway	Any Anomaly/Agent	Mutated	Amplified	Deleted	Agents
ADP-Ribosylation Factor	●		●		
AKAP95 Role in Mitosis and Chromosome Dynamics					
AKT Signaling Pathway					
ALK in Cardiac Myocytes	●				●
ATM Signaling Pathway	●	●			●
Acetylation and Deacetylation of RelA in Nucleus					
Actions of Nitric Oxide in the Heart					
Activation of CSK Inhibits Signaling through the T Cell Receptor	●		●		●
Activation of PKC through G-Protein Coupled Receptors					
Activation of cAMP-dependent Protein Kinase, PKA	●		●		
Acute Myocardial Infarction	●			●	
Adhesion Molecules on Lymphocyte					
Adhesion and Diapedesis of Granulocytes					
Adhesion and Diapedesis of Lymphocytes	●				●
Agrin in Postsynaptic Differentiation	●	●	●	●	
AhR Signal Transduction Pathway					
Alpha-synuclein and Parkin-mediated Proteolysis in Parkinson';s Disease					
Alternative Complement Pathway	●				●
Angiotensin II Mediated Activation of JNK Pathway via Pyk2 Dependent Signaling	●	●	●		

caBIG

EGFR Network Mutation Profile Through CMA



A Systems Approach:
BIG Health Consortium™

"The world we have created today has problems which cannot be solved by thinking the way we thought when we created them."*

**Albert Einstein*

BIG Health Consortium™

Vision:

A biomedical system that synergizes the capabilities of the entire community to realize the promise of personalized medicine

Mission:

The BIG Health Consortium™ is a collaboration among stakeholders in biomedicine, including ***government, academe, industry, non-profit, and consumers,*** who come together in a novel organizational framework ***to demonstrate the feasibility and benefits of the personalized medicine paradigm***.

Strategy:

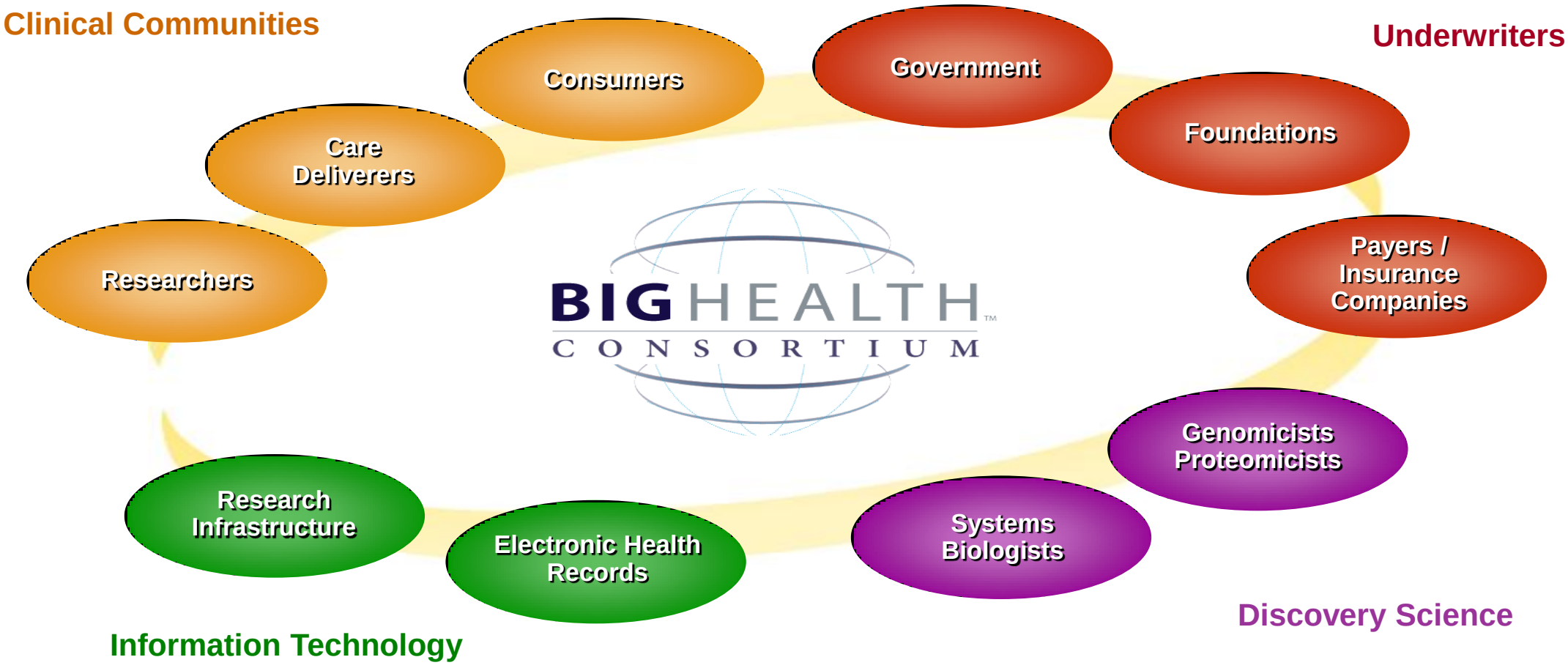
Through a series of personalized medicine **demonstration projects**, with an expanding number of collaborators, BIG Health will **bootstrap** a new approach in which clinical care, clinical research, and scientific discovery are linked.

BIG Health Goals

BIG Health will demonstrate that:

- Loosely-coupled sectors within life sciences and health care can come together in an ecosystem to implement personalized medicine real-world projects, in real time.
- The tools, infrastructure and standards of NCI's informatics infrastructure (caBIG®) can be applied to linking this ecosystem.
- Such an ecosystem can be financially self-sustaining.
- ***Clinical care, clinical research, and scientific discovery can be connected in a seamless continuum that speeds innovation and benefits patients.***

BIG Health Ecosystem



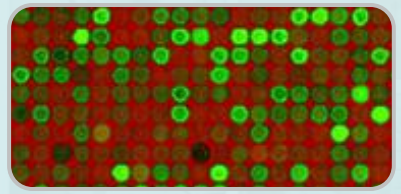
Research

Participants

Patients join research networks, grant consent, agree to be “sought” and to enroll – “on-demand” participants

Biospecimen Collections

Researchers can access and query large collections of well-characterized, clinically annotated specimens



Discovery of Correlations

Biomarkers are identified and validated; disease sub-groups emerge

BCM
Baylor College of Medicine

Individualization of Treatment

Patients are identified by sub-groups and treated appropriately

Clinical Practice

Electronic Health Records

EHRs can connect to clinical trials in hospital settings

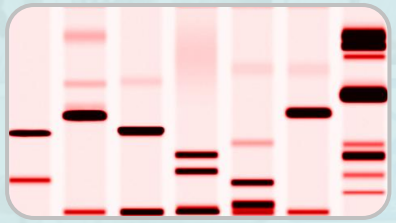
Research Finding Knowledgebases

Large-scale databases of latest research findings are connected to health delivery encounter

Learning Healthcare System

Local and national clinical encounter information is fed back to care providers to help inform clinical decision making

Consumer



My Genomic Profile

Consumers get their genetic and predisposition risk information

My Prevention Strategies

Consumers work with genetic counselors; coordinate with health care provider

My Clinical Record

Consumers link to their clinical histories with genetic profiles; access clinical research; participate in volunteer networks

BIG HEALTH CONSORTIUM

Standards

Interoperability

Data Sharing

caGRID

Connectivity

BIG HEALTH CONSORTIUM

BIG Health Ecosystem

Academic/Health Care Delivery

- Baylor College of Medicine
- Duke
- Georgetown
- Institute of Medicine
- NCCCP
- UCSF

Diagnostic

- Cellpoint
- Genzyme Genetics
- Monogram Biosciences

(Bio)Pharmaceutical

- Novartis
- Exelixis
- Genzyme
- J&J

IT/EHR/PHR

- Cerner
- Health IT Inc.
- Microsoft
- Oracle
- IBM

Foundations/Non-Profit/Advocacy

- Brookings Institution
- Canyon Ranch Institute
- CollabRx
- Critical Path Institute
- *FasterCures*
- Lance Armstrong Foundation
- Personalized Medicine Coalition

Health Care Consultancy

- Booz Allen Hamilton
- Feinstein Kean Healthcare
- Deliotte

Government

- CDC
- HHS Personalized Health Care Initiative
- NCI
- ONC

Payers

- Kaiser Permanente

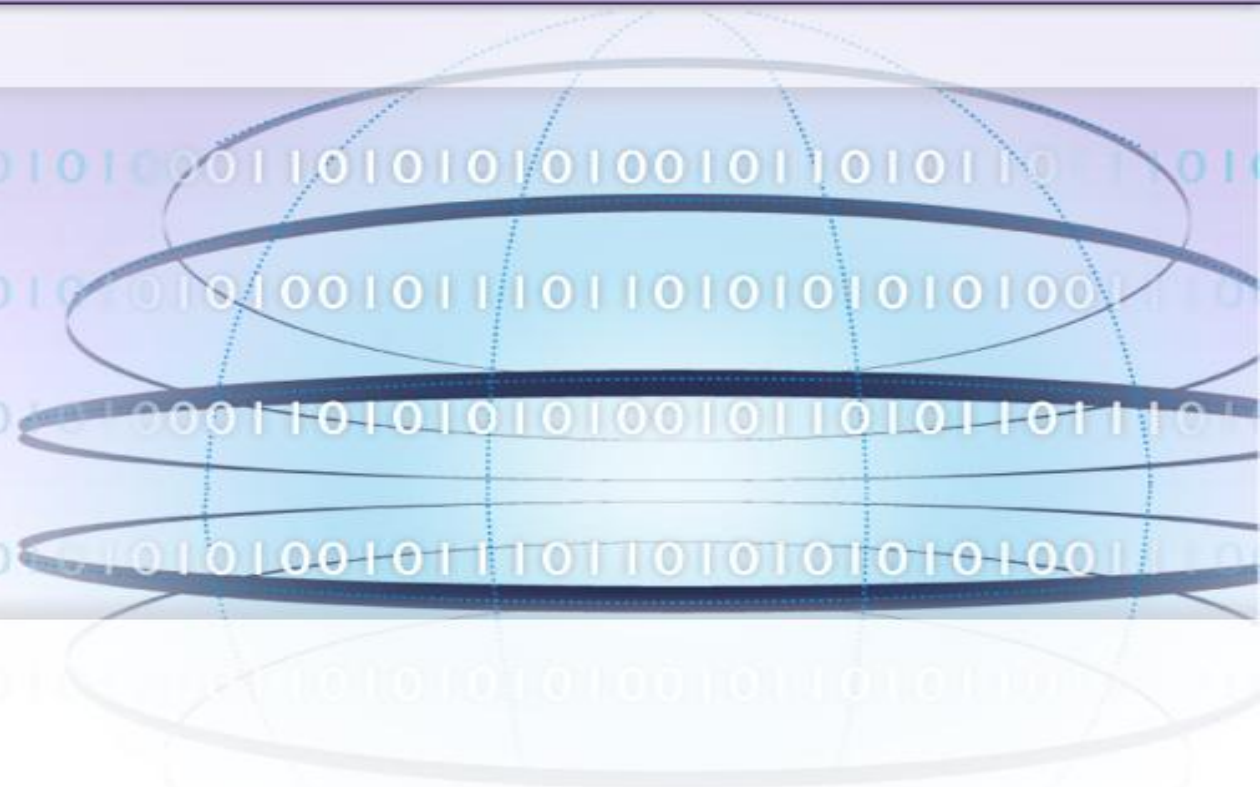
Venture Capital

- Health Evolution Partners
- MDV

Personal Genomics

- Navigenics
- 23 and Me

BIG HEALTH
CONSORTIUM™



http://BIGHealthConsortium.org

Join the effort!!!

More information:

caBIG.cancer.gov

Join caBIG effort:

caBIG.nci.nih.gov

caBIG®: Power of Connection