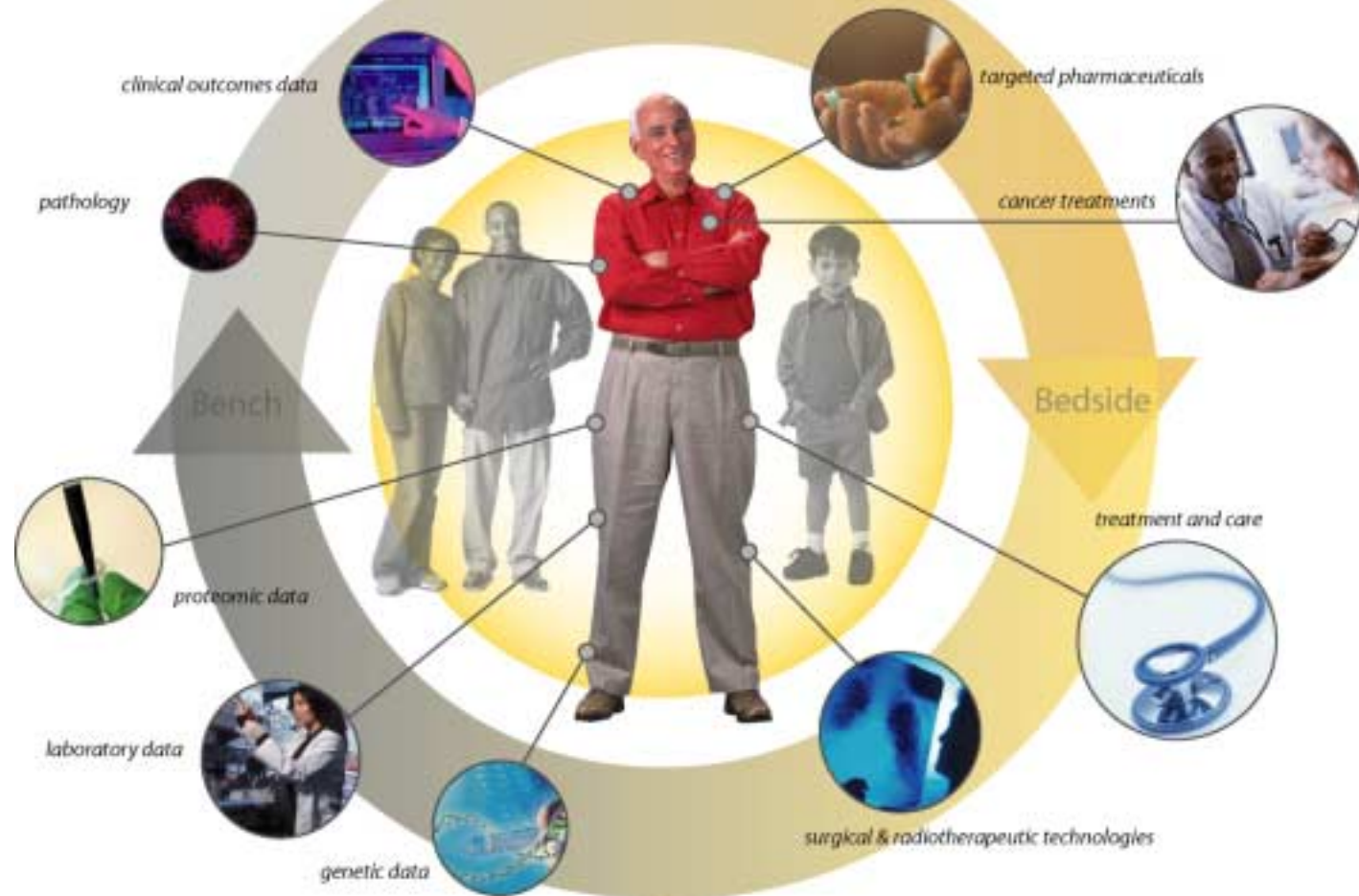# caBIG's SOA Case Study:
Service Quality, Delivery, and Efficiency

Ken Buetow, Ph.D.
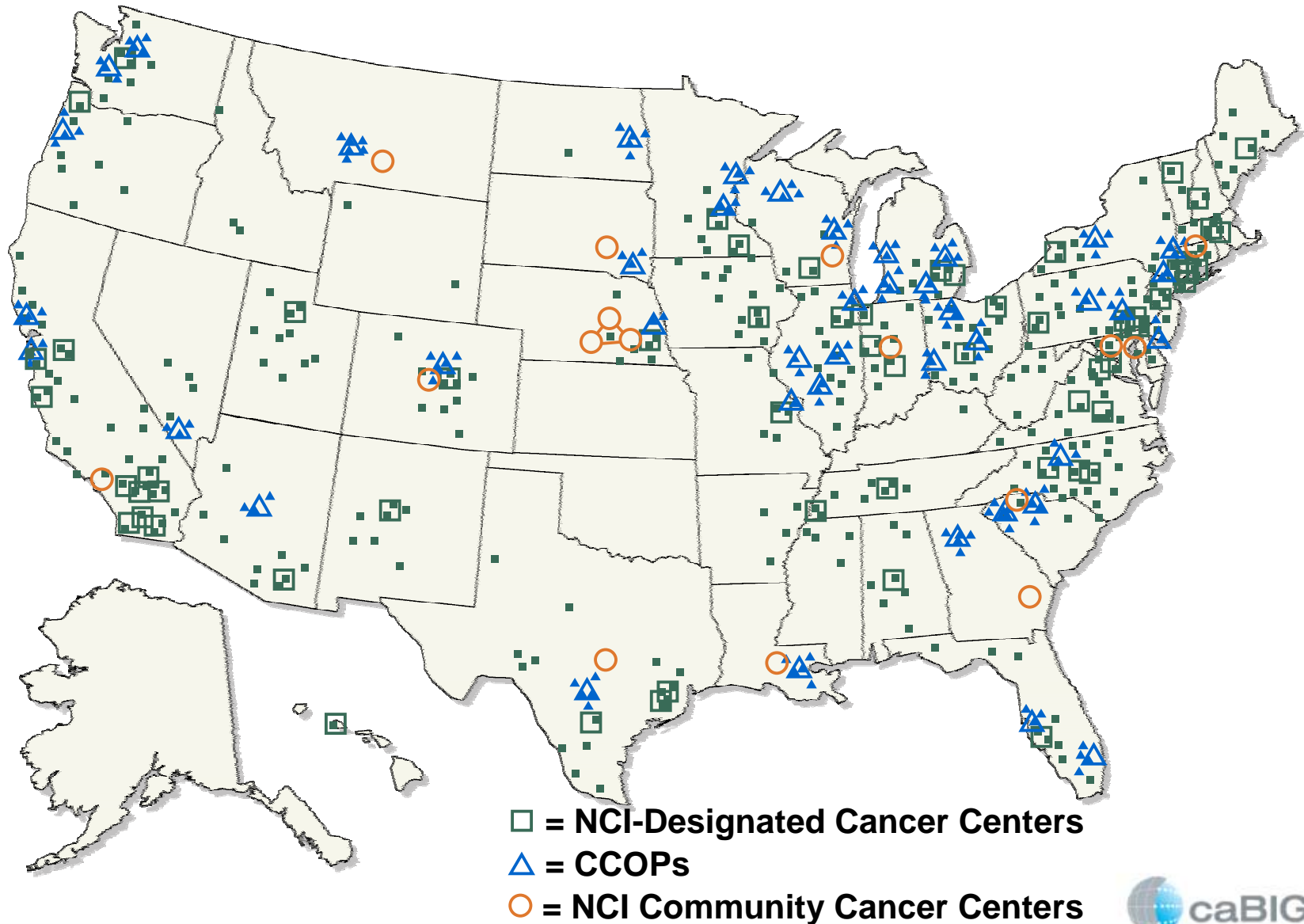
NCI Associate Director

Biomedical Informatics and Information Technology

Individualized, Targeted Cancer Care

clinical outcomes data

pathology

targeted pharmaceuticals

cancer treatments

Bench

Bedside

proteomic data

laboratory data

genetic data

treatment and care

surgical & radiotherapeutic technologies

caBIG™ cancer Biomedical Informatics Grid™

NCI Cancer Centers Programs

□ = NCI-Designated Cancer Centers
△ = CCOPs
○ = NCI Community Cancer Centers

Integrated Health Infrastructure

Clinical Systems

Labs, EMR, Tissue, etc.

Patient Health Record

HL7/CAM SDK

HL7-v2.x, other

HL7-v3

Clinical Research

etc.
Lifecycle Management
Adverse Events
Participant Registry
EDC
Clinical Data Mgmt

De-identification Services

Research Data Warehouse

BRIDG, HL7-v3, Janus, CDUS, Theradex

Clinical Research Information Exchange

FIReBIRD **Janus**

CDS

User Tools

BRIDG, HL7-v3, Janus

External Reporting

FDA
SPONSOR
NCI
other

caBIG cancer biomedical informatics Grid

THE CANCER GENOME ATLAS

The Cancer Genome Atlas (TCGA)

**A 3-year pilot project of the
National Cancer Institute (NCI) and
the National Human Genome Research Institute (NHGRI)
to increase our comprehensive understanding of the
genetic basis of cancer**

**It is anticipated that TCGA's
integrated database of molecular and clinical information
will provide scientists unprecedented opportunities to discover
and develop a new generation of targeted diagnostics, therapies,
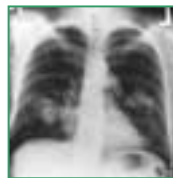and preventives for cancer.**

caBIG

THE CANCER GENOME ATLAS

Challenge One: Connecting multiple sources, experiments, and data types

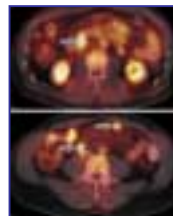## Three forms of cancer

Multiple data types

glioblastoma multiforme
(brain)
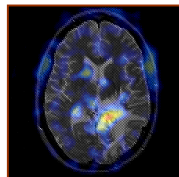
squamous carcinoma
(lung)

serous
cystadenocarcinoma
(ovarian)

12 Organizations

Biospecimen Core
Resource

7 Cancer Genomic
Characterization
Centers

3 Genome
Sequencing
Centers

Data Coordinating
Center

- 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

caBIG

THE CANCER GENOME ATLAS
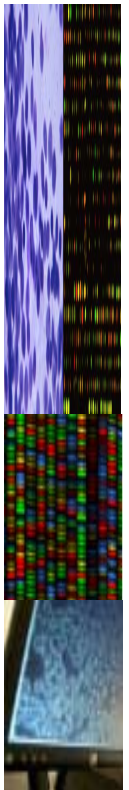
Challenge Two: Empowering researchers to query increasingly complex layers of cancer biology, from gene to clinical phenotype, as a whole

Gene ◄► Genome ◄► Genomes ◄► Pathways ◄► Clinical Outcomes

All from their computer

caBIG

A necessary Bridge Between Research and Care Delivery

clinical practice

Platforms
- Medical centers
- Community hospitals
- Private practice
- Government

Shared –
Health Information Technology
- Infrastructure
- Standards
- Development

E-health record

Platforms
- Academic centers
- Pharma/CROs
- Biotech
- Government

clinical research

Practice outcomes

Extended participant access

Molecular medicine

Trials outcomes

Efficient Care, Better Health

Extended –
Next Generation Molecular Medicine
- Infrastructure
- Standards
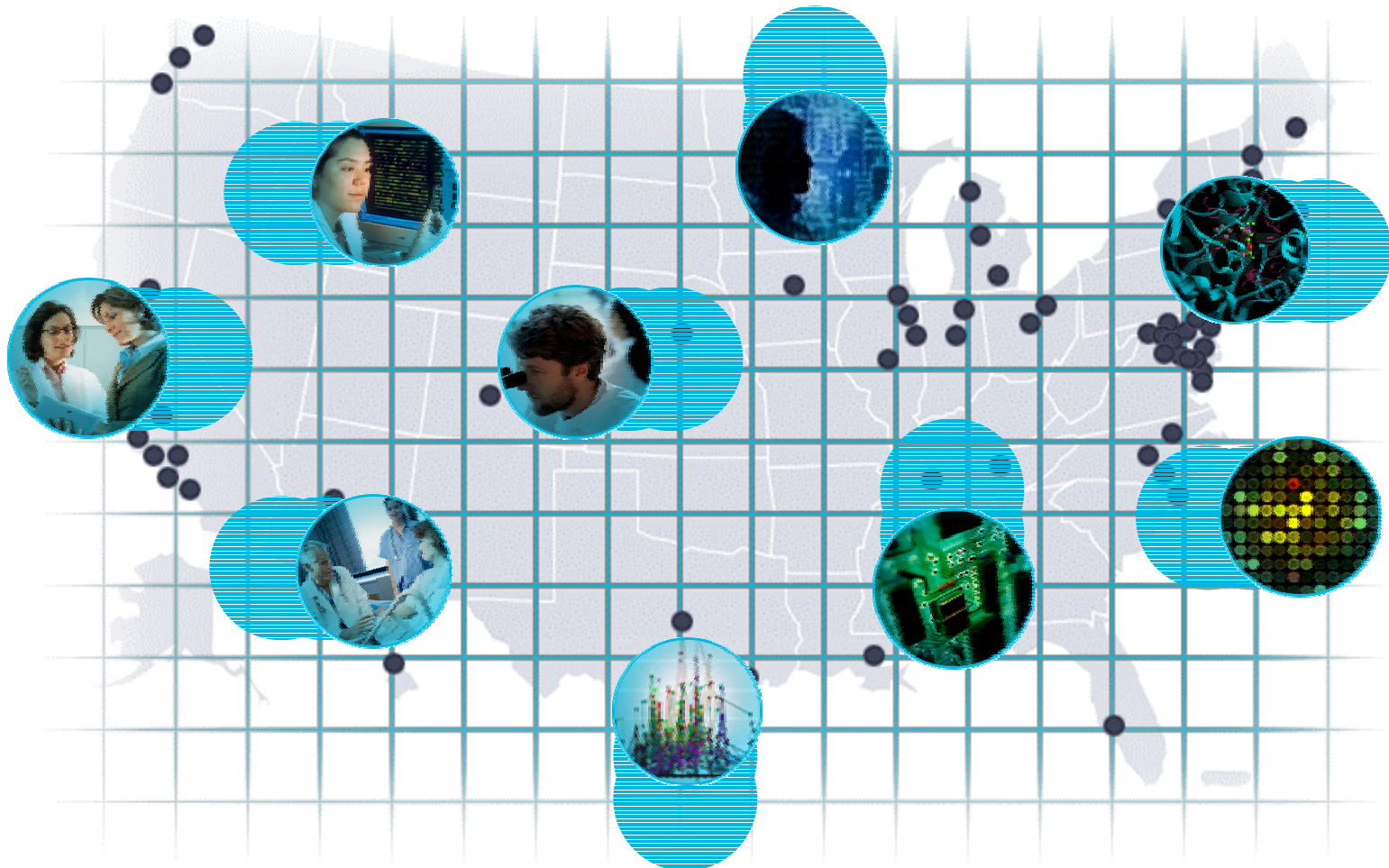- Development

caBIG

The caBIG™ Initiative

caBIG™ Goal
A virtual network of interconnected data, individuals, and organizations that whose goal is to redefine how research is conducted, care is provided, and patients/participants interact with the biomedical research enterprise.

caBIG™ Vision

- **Connect** the cancer research community through a shareable, interoperable electronic 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™ Collaboration: Academic, Government, Industry

Alabama

Iowa

New York

Indianapolis:
Indiana University Cancer Center
Regenstrief Institute, Inc.

Lebanon:
Dartmouth College
Dartmouth-Hitchcock Medical Center

Fred Hutchinson Cancer Research Center
International
Paris, France:
Sanofi Aventis

caBIG Community

Clinical Trial Management Systems	Addresses the need for consistent, open and comprehensive tools for clinical trials management.
Integrative Cancer Research	**Provides tools and systems to enable integration and sharing of information.**
Tissue Banks & Pathology Tools	**Provides for the integration, development, and implementation of tissue and pathology tools.**
In vivo **Imaging**	**Provides for the sharing and analysis of in vivo imaging data.**

Responsible for evaluating, developing, and integrating systems for vocabulary and ontology content, standards, and software systems for content delivery.	**Vocabularies & Common Data Elements**
Developing architectural standards and architecture necessary for other workspaces.	**Architecture**

Data Sharing and Intellectual Capital	sharing of data, applications and infrastructure within the cancer community.
Training	**training in the use of the caBIG™ resources including on-line turtorials, workshops, training programs.**
Strategic Planning	**Assists in identifying strategic priorities for the development and evolution of the caBIG™ effort.**

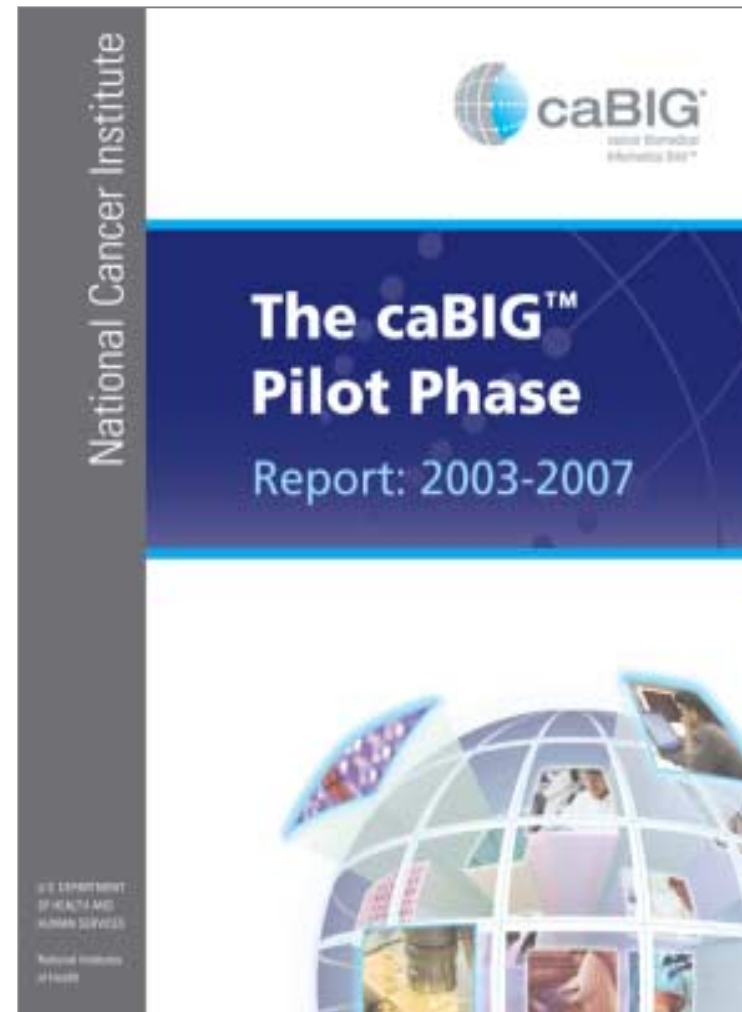caBIG cancer biomedical informatics Grid™

Herding Cats…

The caBIG™ Current Snapshot

- **caBIG in a nutshell**
  - 190 participating organizations
  - 300 software components
  - 40+ end-user applications in discovery, clinical trials management, biospecimen management, etc.
  - caGrid providing data transmission network that "connects" everyone

- **43 NCI-designated Cancer Centers actively participating in caBIG™ deployment program**
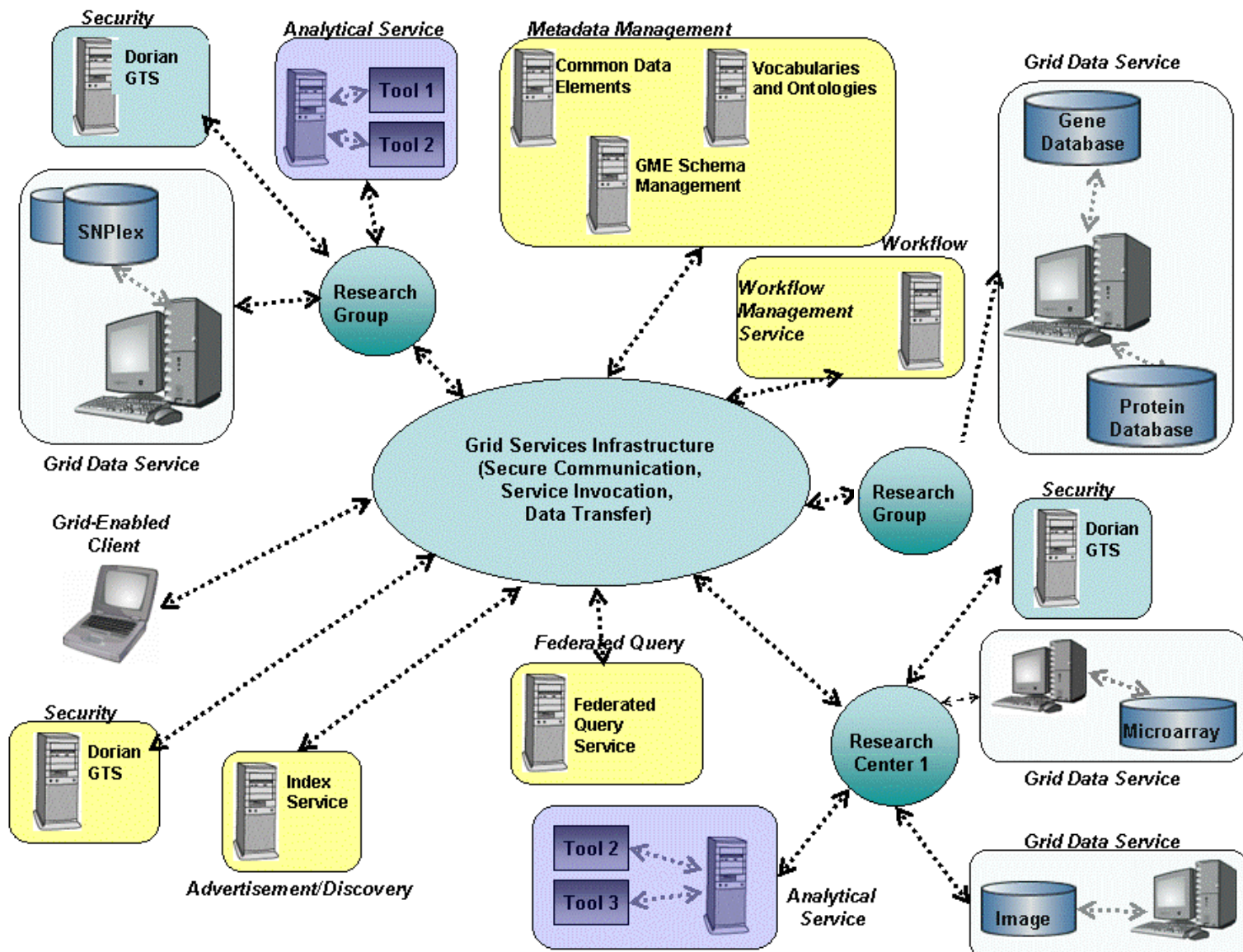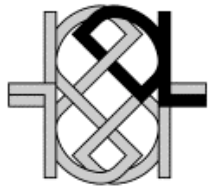- **16 NCI Community Cancer Centers connecting through caBIG**

National Cancer Institute

caBIG·

The caBIG™ Pilot Phase

Report: 2003-2007

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

caBIG·

caBIG: an open SOA with shared community semantics

caGrid 1.1 Conceptual View

Industry Standards

Description Logic

Clinical Data Interchange

ISO/IEC 11179 Metadata Registries
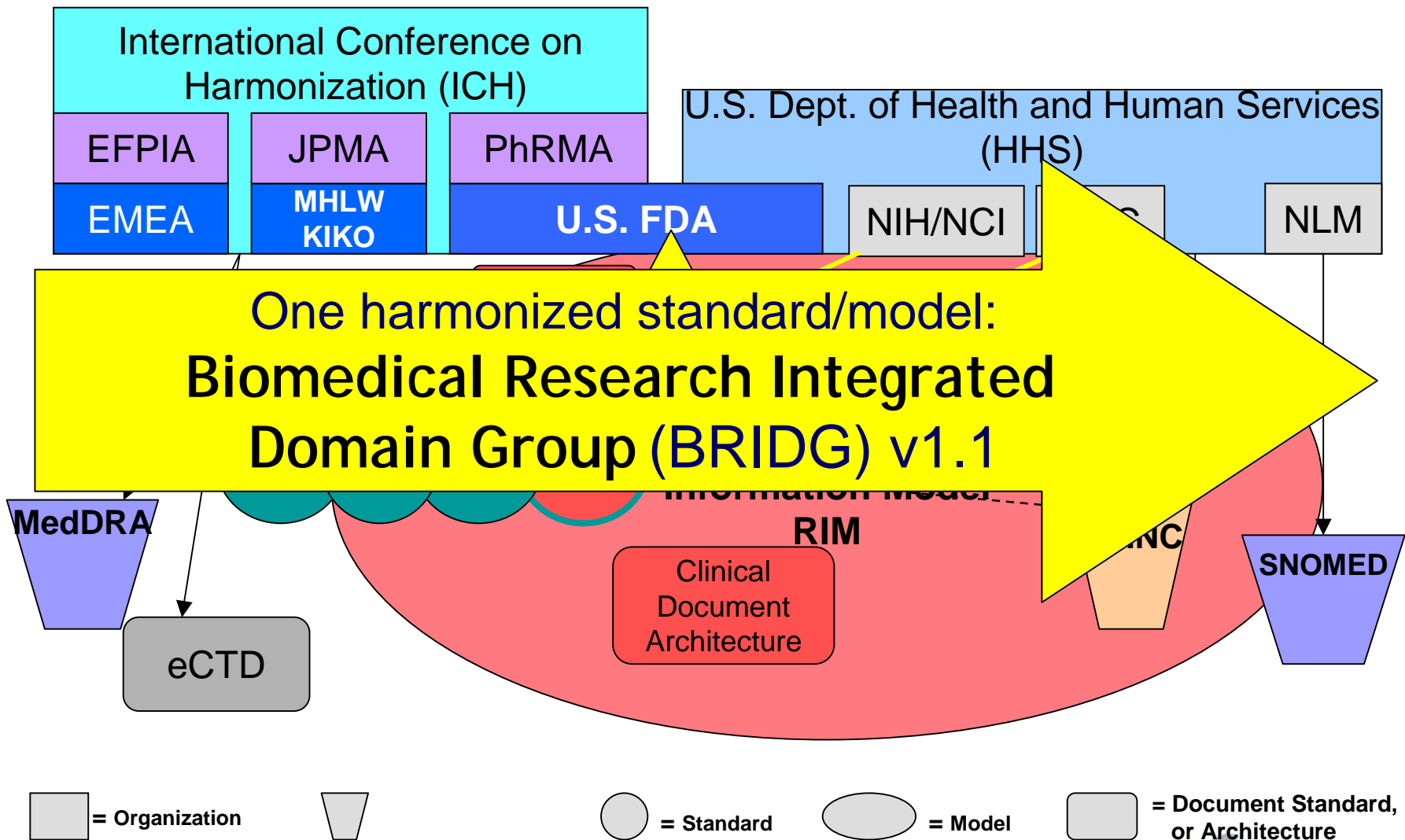
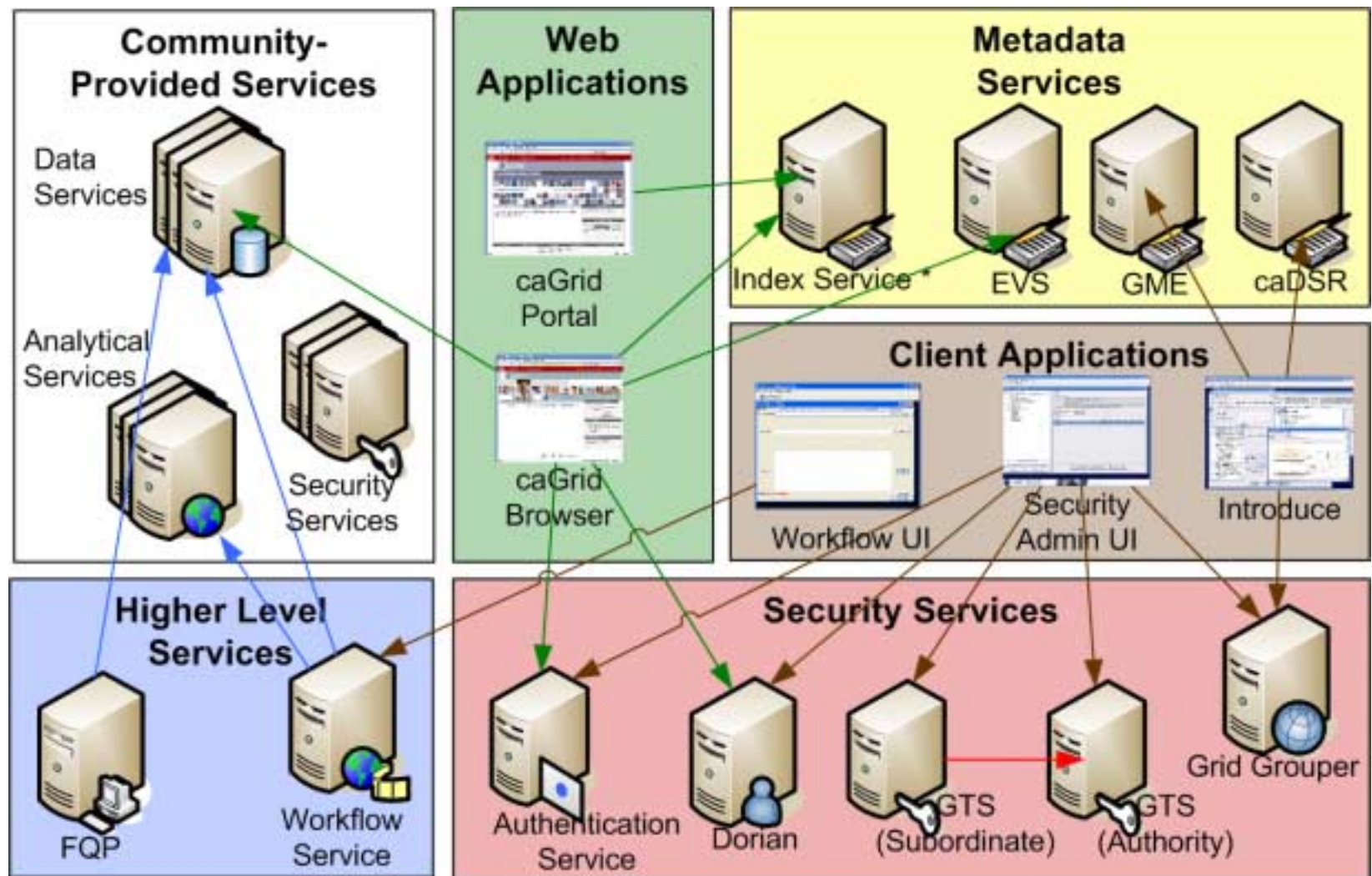UNIFIED MODELING LANGUAGE™

XML Schema Web Services

Web Services Resource Framework (WSRF)

Systems Interoperability & Harmonization

International Conference on Harmonization (ICH)

EFPIA

JPMA

PhRMA

EMEA

MHLW KIKO

U.S. FDA

U.S. Dept. of Health and Human Services (HHS)

NIH/NCI

NLM

One harmonized standard/model:
Biomedical Research Integrated Domain Group (BRIDG) v1.1

MedDRA

Information Model RIM

Clinical Document Architecture

SNOMED

eCTD

= Organization   = Standard   = Model   = Document Standard, or Architecture

caBIG

caGrid: caBIG's SOA Environment

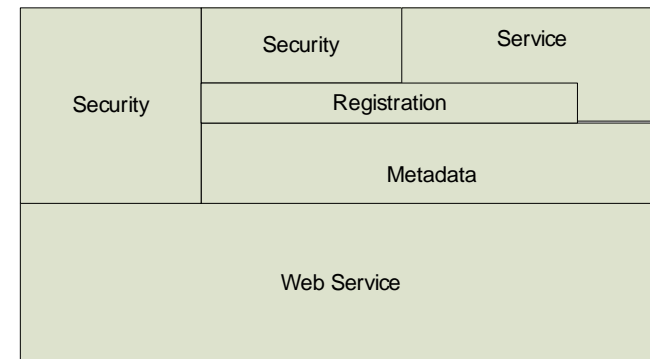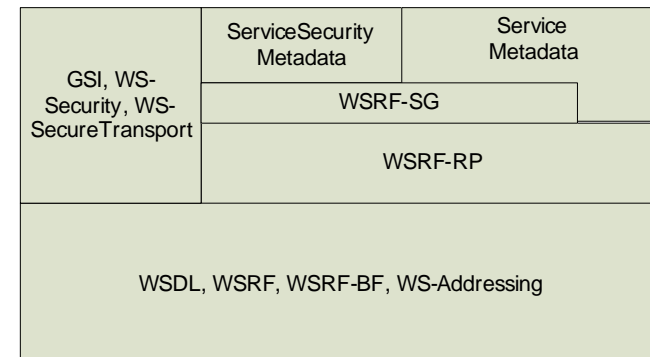*All Services Register with the Index Service

Standards Based

- **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

Security	Security	Service
	Registration	
	Metadata	
Web Service		

Layer Specifications

GSI, WS-Security, WS-SecureTransport	ServiceSecurity Metadata	Service Metadata
	WSRF-SG	
	WSRF-RP	
WSDL, WSRF, WSRF-BF, WS-Addressing		

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

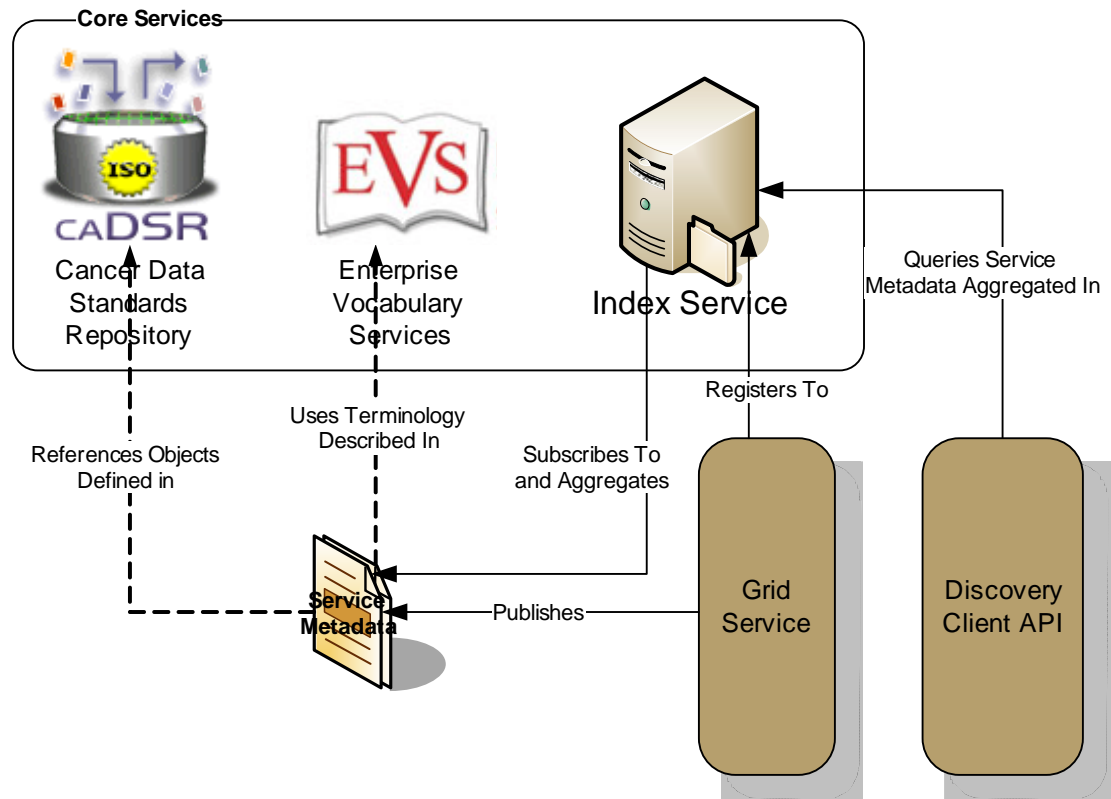**Service Implementation**

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

Core Services

Cancer Data Standards Repository

Enterprise Vocabulary Services

Index Service

Queries Service Metadata Aggregated In

Uses Terminology Described In

References Objects Defined in

Registers To

Subscribes To and Aggregates

Service Metadata

Publishes

Grid Service

Discovery Client API

caBIG

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.

caBIG

Grid of Grids…
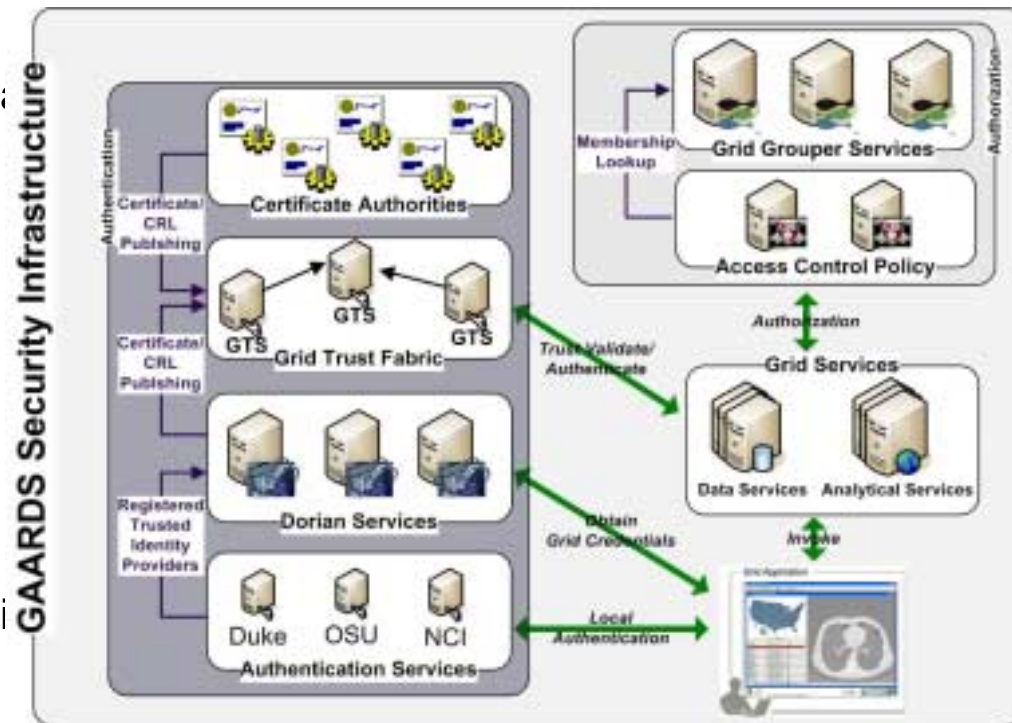
Embracing Enterprise Architecture

- **Enterprise Architecture is essential because…**
  - The Problem Space (translational medicine) is complex
    - Multiple disciplines and stakeholders
    - Historical isolation
    - Fundamental commitment to change ("We do discovery")
  - The Solution Space is complex
    - Federated architecture
    - Need for of significant computational resources
    - Requirement for shared information semantics

caBIG

Enterprise Architecture:
Our Working Choice (RM-ODP)

- **Reference Model for Open Distributed Processing**
  - ISO Standard (RM – ODP, ISO/IEC IS 10746 | ITU-T X.900)
  - Two Components
    - Five *non-hierarchical* Viewpoints for specifying a Composite Architecture
      - Enterprise (business context)
      - Information (dynamic and static semantics)
      - Computational (logical architecture/design)
      - Engineering (distribution of engineering resources)
      - Technology (technology-specific bindings)
    - Ontology
      - Facilitates the communication between and among viewpoint stakeholders
      - Specifies a common language for Conformance Testing

Join the effort!!!

More information:
caBIG.cancer.gov

Join caBIG effort:
caBIG.nci.nih.gov