

Privacy and Security in a Federated Research Network

Dan Steinberg, JD
HIPAA Summit XIII
Washington, DC
September 26, 2006

Overview

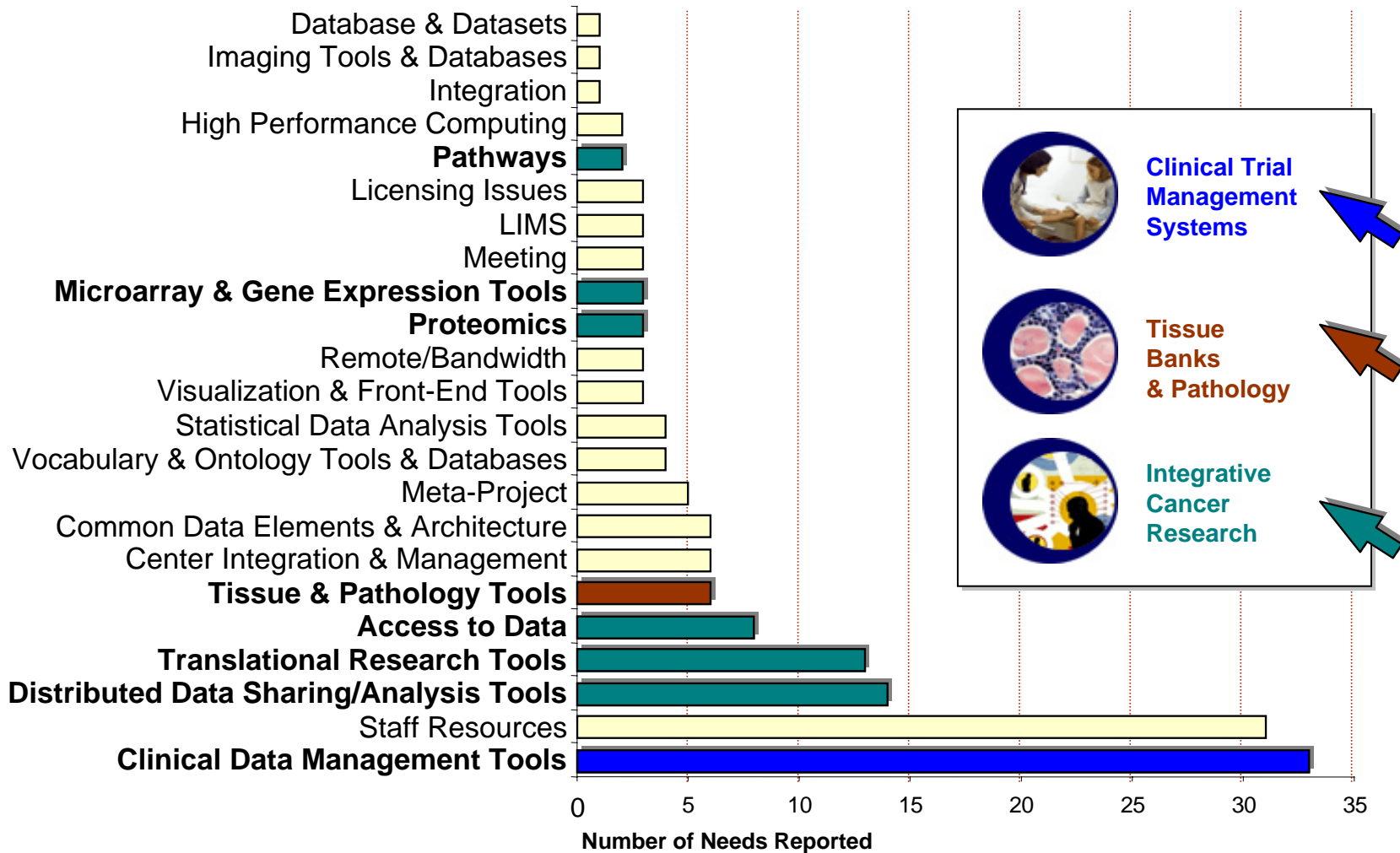
- ▶ **What is caBIG™?**
- ▶ caBIG™ Structure and Oversight
- ▶ HIPAA-Related Issues
- ▶ Issues Identified and Addressed
- ▶ For More Information

caBIG™ is the “World Wide Web of cancer research.”



- The cancer Biomedical Informatics Grid (caBIG™) is a voluntary, virtual network
- caBIG™:
 - Connects individuals and institutions
 - Enables the sharing of data and tools
- The use of caBIG™ resources will speed the delivery of innovative approaches for the prevention and treatment of cancer
- caBIG™ is being developed under the leadership of the National Cancer Institute's Center for Bioinformatics (NCICB)

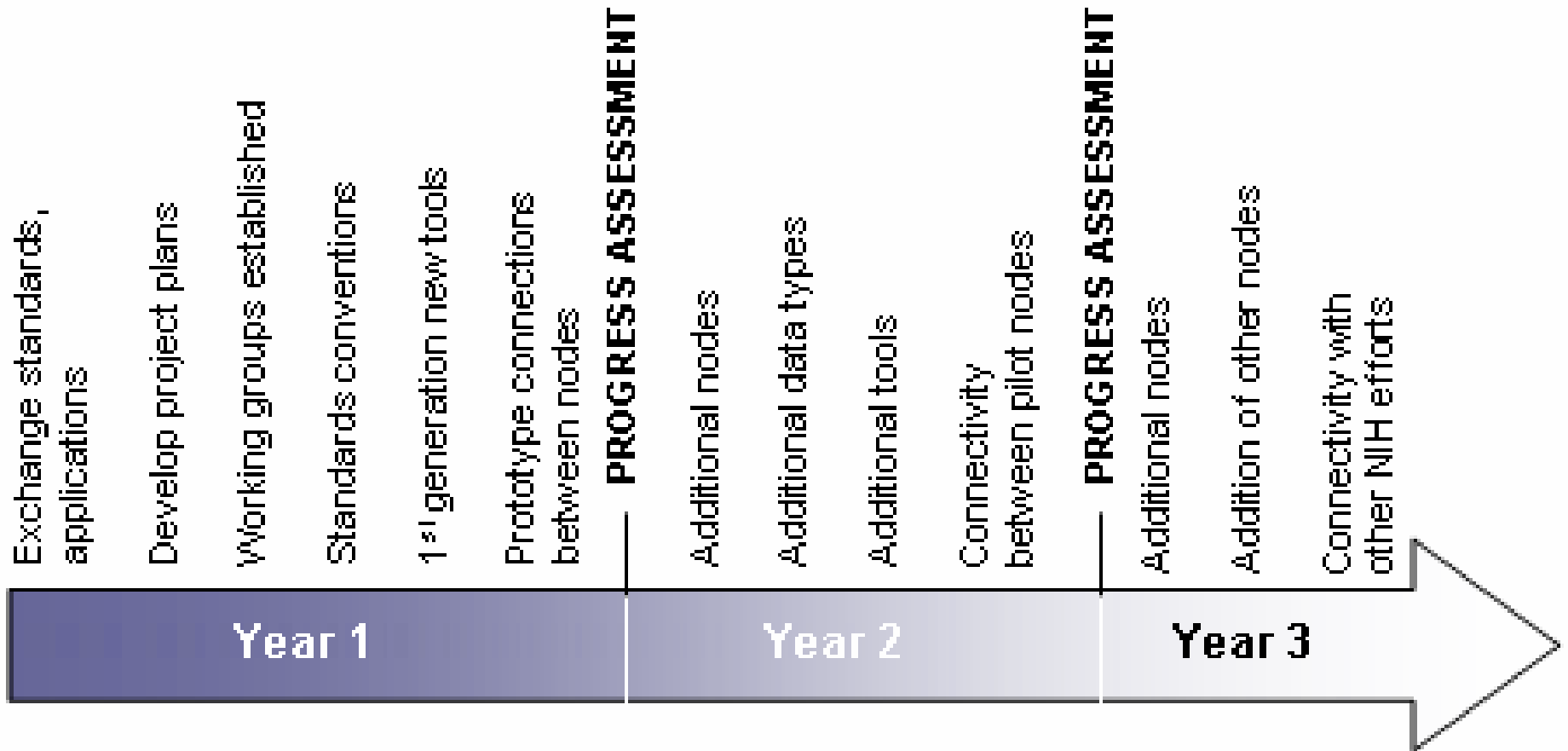
caBIG™ was developed in response to a survey of Cancer Centers' needs.



caBIG™ offers institutions resources to develop tools on the condition that those tools are shared with other institutions that need them.

- ▶ caBIG™ established a pilot network of NCI Cancer Centers
 - Groups agreeing to caBIG™ principles
 - Mixture of capabilities
 - Mixture of contributions
- ▶ We have expanded the number and types of participants
- ▶ We have also established a consortium development process
 - Collect and share expertise
 - Identify and prioritize community needs
 - Expand development efforts

The focus of programmatic activities has shifted as the caBIG™ initiative has matured over the life of the pilot.



The caBIG™ community now includes hundreds of participants at dozens of institutions.

- ▶ Many developers and adopters are NCI designated cancer centers
 - Designation based on competitive, peer-reviewed grant application among US cancer research institutions
 - Virtually all (approximately 50) participate in caBIG™ as developers, adopters, or participants
- ▶ Booz Allen is the prime contractor, providing management, coordination, and other supervisory activities
- ▶ Other participants include:
 - Organizations in the public and private sectors, including nonprofits
 - Scientists
 - Informaticists
 - Patient advocates
 - Commercial and industry groups (e.g., information technology companies, software vendors, pharmaceutical companies, biotechnology companies)
- ▶ Participation is open to any interested party, including those beyond the cancer research community

Overview

- ▶ What is caBIG™?
- ▶ **caBIG™ Structure and Oversight**
- ▶ HIPAA-Related Issues
- ▶ Issues Identified and Addressed
- ▶ For More Information

The work of caBIG™ is guided by four fundamental principles.

▶ Open source

- Developers of software tools and applications funded by NCI through caBIG™ must make the source code publicly available [in “non-viral” terms]

▶ Open access

- caBIG™ resources must be freely obtainable

▶ Open development

- caBIG™ products are developed through an open, participatory process, including open participation in regularly scheduled teleconferences and periodic face-to-face meetings

▶ Federation

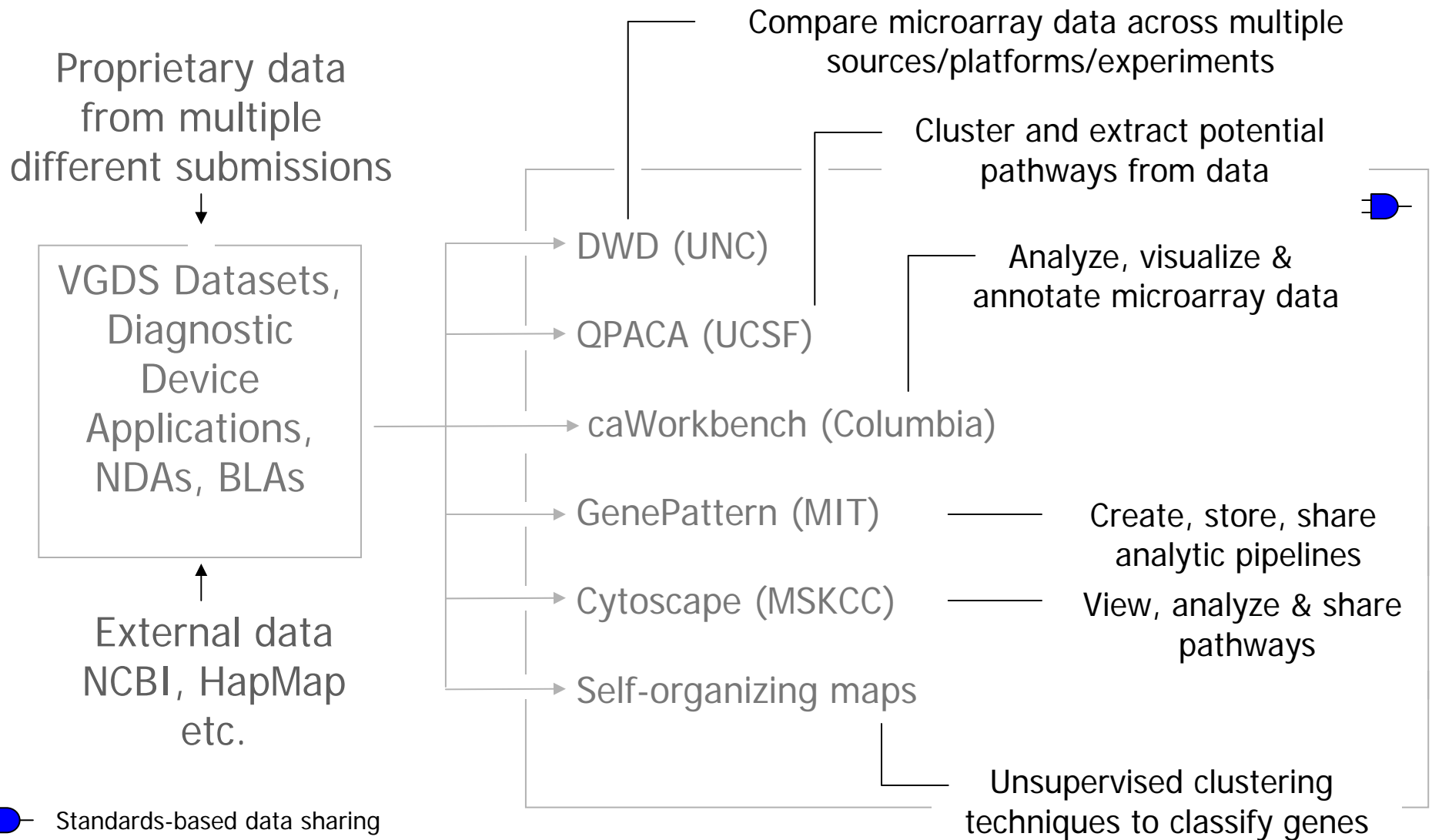
- caBIG™ is designed to be a network of systems that can be locally controlled
- This is significant for individual institutions with obligations under HIPAA because data is retained locally

The work of caBIG™ is divided by function among “workspaces.”

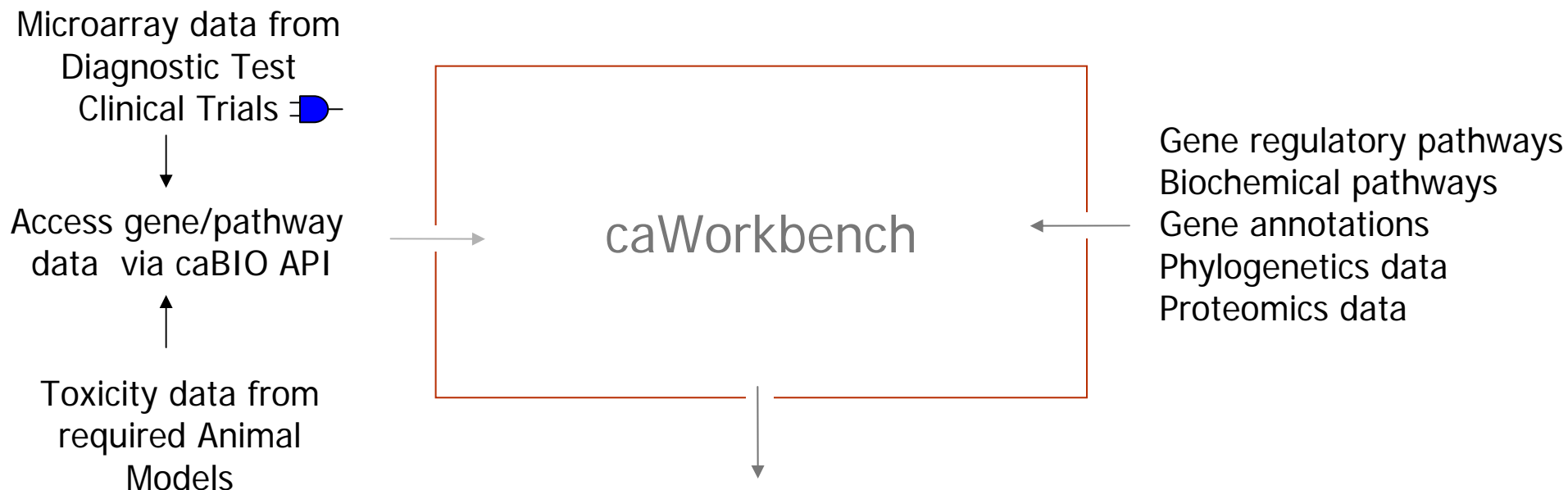
- ▶ **Domain Workspaces:** Develop software tools designed to address research needs
 - Integrative Cancer Research (ICR)
 - Tissue Banks and Pathology Tools (TBPT)
 - Clinical Trials Management Systems (CTMS)
 - Imaging (IMAG)
- ▶ **Cross Cutting Workspaces:** Develop the infrastructure necessary to integrate the tools developed by the domain workspaces
 - Vocabularies and Common Data Elements (VCDE)
 - Architecture (ARCH)
- ▶ **Strategic Level Workspaces:** Provide policy, governance, and advisory functions
 - Training and Documentation
 - Strategic Planning (SP)
 - Data Sharing and Intellectual Capital (DSIC)

Examples of caBIG™ Tools

Examples of compatible Gene Expression Data Analysis Tools



Example - Analysis & Visualization of Pathway Data: caWorkbench




Align sequences from different Drug Applications with BLAST searches

Apply pattern discovery algorithms

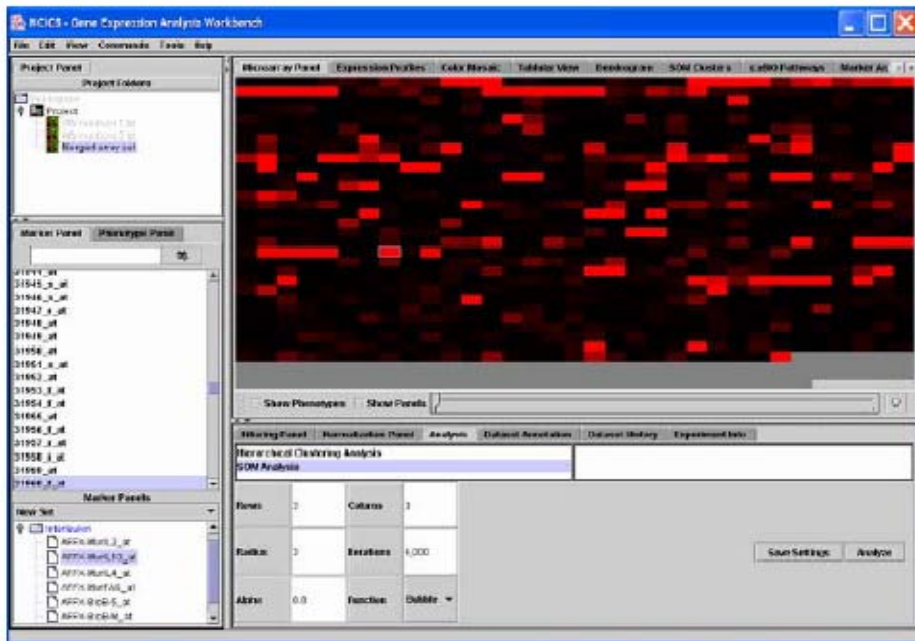
Display EST mapping with genome sequence retrieval

Localize gene expression in tissues and organ with CGAP microarray data,

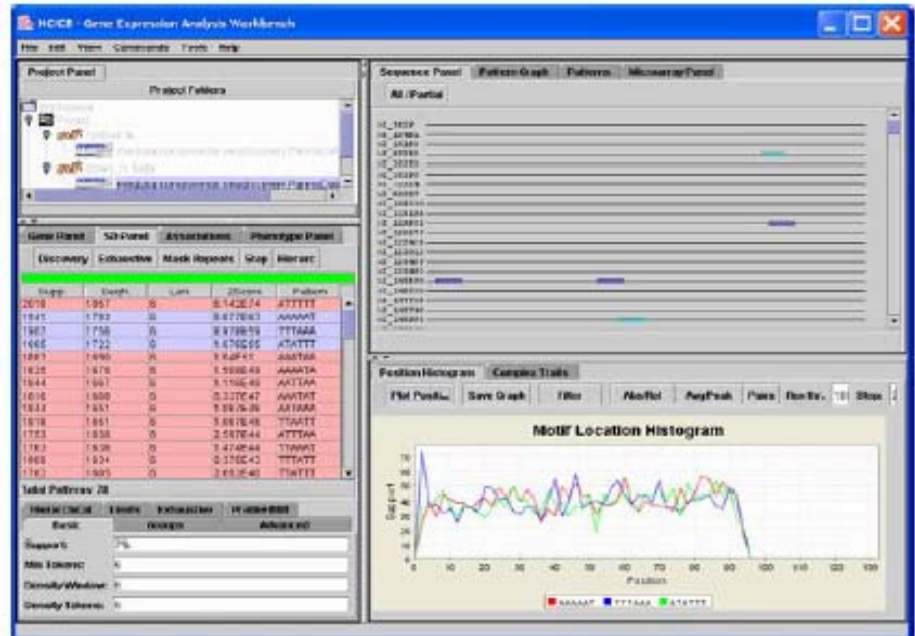
Map drug interaction on Pathway diagrams

 Standards-based data sharing

caWorkbench



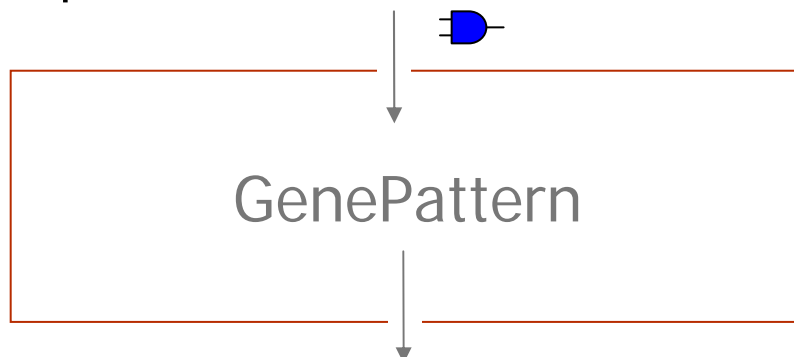
A fully populated framework for Microarray Analysis



Sequence Pattern Discovery Application

Example - Creation & Sharing of Analytic Pipelines: GenePattern

Gene expression data from Clinical Trials



Analyze standard global sequence

Supervised and unsupervised learning


Select genes that most closely resemble a profile

Select genes that most closely resemble a continuous profile

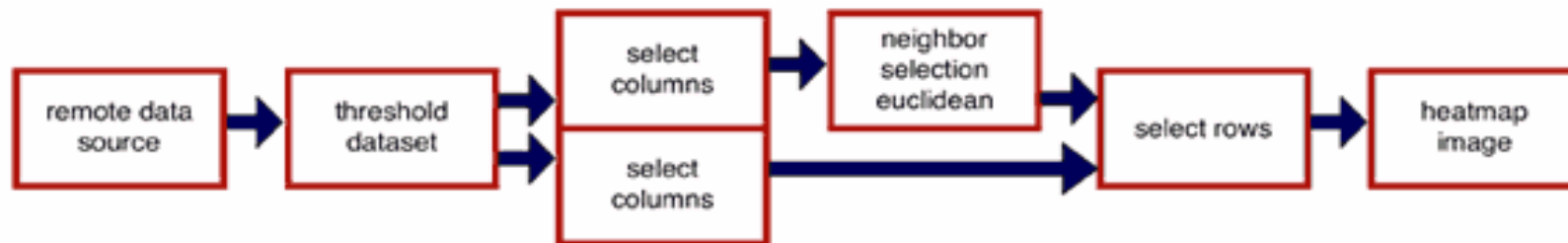
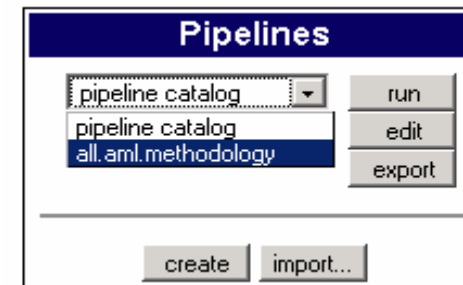
Creates a heat map graphic from a dataset

Visualize clusters

Chain tasks together to create, encapsulate, reproduce, and share methodologies

 Standards-based data sharing

GenePattern



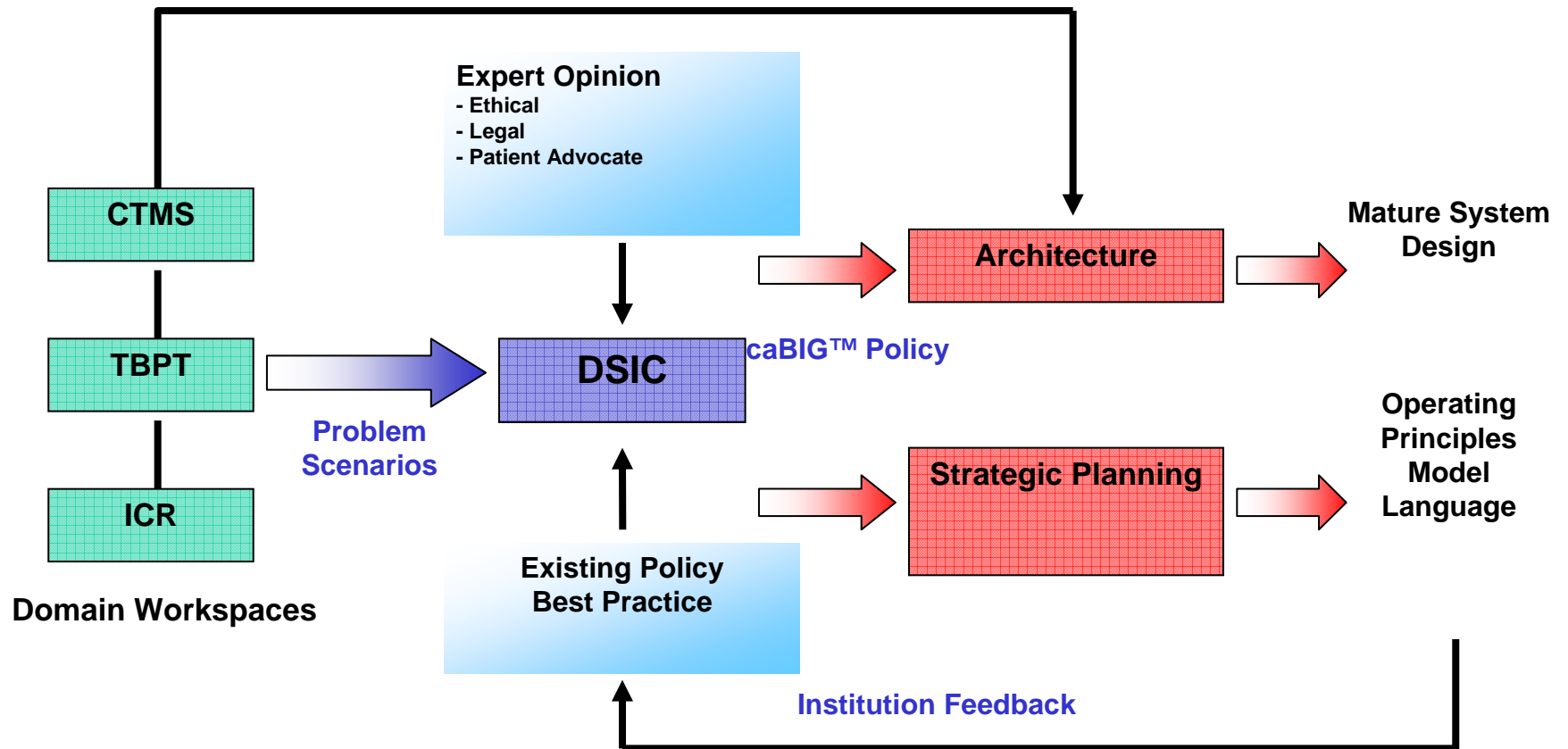
Overview

- ▶ What is caBIG™?
- ▶ caBIG™ Structure and Oversight
- ▶ **HIPAA-Related Issues**
- ▶ Issues Identified and Addressed
- ▶ For More Information

Issues related to privacy and security are addressed in the Data Sharing and Intellectual Capital Workspace (DSIC WS).

- ▶ Goal: identify and then propose solutions to potential barriers to data and resource sharing and other collaborative work across the caBIG community
- ▶ These barriers may arise from law, regulation, institutional policies and desire to protect intellectual property interests
- ▶ DSIC WS contains about twenty regular participants, and an additional twenty to thirty ad hoc participants, with a wide range of perspectives and expertise
- ▶ Legal and policy requirements related to privacy and security drivers include
 - HIPAA Privacy Rule
 - HIPAA Security Rule
 - The Common Rule for Human Subjects Research
 - FDA Regulations on Human Subjects
 - 21 CFR Part 11
 - State and institutional requirements.

DSIC WS receives concerns from the community and develops responses.



Responsibility for HIPAA compliance remains with caBIG participants.

- ▶ caBIG™'s federated structure allows data to reside on the servers of its originator or owner
- ▶ Participants maintain control and responsibility for the data
- ▶ Data that is shared must be either de-identified or shared pursuant to HIPAA-compliant agreements between providers and recipients of the data
- ▶ Data use provisions in the funding agreements with caBIG participants allocate responsibility for compliance with HIPAA to the institutions of the caBIG™ participants:
 - *“Any Data or related information delivered or made available to [NCI either directly or via the prime contractor], the caBIG™ Community or the public pursuant to Task Orders and published by Subcontractor shall be published in accordance with institutional review board (IRB) requirements, state privacy laws, and the HIPAA Privacy Rule.”*

Similar requirements are reflected in task orders issuing from the prime contractor for development, adoption, or other research projects.

- ▶ **Sample language addressing “Data Use, Disclosure of Information and Handling of Sensitive Information” (for a software development project):**

The developer must address the potential sensitivity of the information collected, information security issues, local Institutional Review Board (IRB) requirements and the Health Insurance Portability and Accountability Act (HIPAA) of 1996 in its design of the system in question. The system must accommodate the needs and actual uses, related to these laws and regulations, of caBIG™ participants.

Final regulations issued by the Department of Health and Human Services (DHHS) provide privacy and security standards that must be observed in the handling of patient data resulting from biomedical research. HIPAA privacy standards will be used to establish safeguards and restrictions for the use and disclosure of research records. HIPAA security standards will be used to help Cancer Centers implement administrative, physical, and technical safeguards to protect electronic health information. Improper use or disclosure of sensitive information under the rules may be subject to criminal or civil sanctions prescribed in HIPAA.

DSIC WS seeks to promote data sharing through caBIG™ by identifying barriers and proposing solutions.

- ▶ Policies — For adherence to requirements of participation in the caBIG™ community
 - Licensing terms
 - Disclosure of conflicts of interest
 - Standards of review for deliverables
- ▶ Guidelines — Recommended practices to enable data sharing
 - Based on practices of other large data sharing initiatives and proposals to cover gaps not yet addressed
 - Identification of standard forms
- ▶ Templates — Recommended language for data sharing documents
 - Authorization and consent forms
 - Disclaimers and notices to be displayed to users of software
 - Limited Data Set agreements
- ▶ Education/training — caBIG™ Annual Meeting, periodic face-to-face meetings, and regular teleconferences.

DSIC WS is compiling a survey of practices in other initiatives with expectations for large scale data sharing to address caBIG™ community needs.

- ▶ Other NCI Sponsored Initiatives
 - Biomedical Informatics Research Network ([BIRN](#))
 - Cooperative Breast Cancer Tissue Resource ([CBCTR](#))
 - Cooperative Human Tissue Network ([CHTN](#))
 - Cooperative Prostate Cancer Tissue Resource ([CPCTR](#))
 - Early Detection Research Network ([EDRN](#))
 - Prostate SPORE National Biospecimen Network ([NBN](#)) Pilot
 - Pennsylvania Cancer Alliance Bioinformatics Consortium ([PCABC](#))
 - Shared Pathology Informatics Network ([SPIN](#))
- ▶ Other Large Scale Biomedical Data Sharing Initiatives
 - Autism Genetic Resource Exchange ([AGRE](#))
 - NIMH Human Genetics Initiative ([HGI](#))
 - Informatics for Integrating Biology and the Bedside ([I2B2](#))
 - NIGMS Protein Structure Initiative ([PSI](#))
 - European Organization for the Research and Treatment of Cancer (EORTC) [Virtual Tumour Bank](#)
- ▶ Other Grid Computing Projects (In progress)
 - DOE Earth Systems Grid ([ESG](#))
 - United Kingdom's [myGrid project](#)
 - National Science Foundation (NSF) Open Science Grid ([OSG](#))
 - DOE Particle Physics Data Grid ([PPDG](#))
 - Department of Energy (DOE) [Science Grid](#)
 - Argonne National Labs' [TeraGrid](#)

Our research thus far has identified many challenges.

- ▶ Restrictions on sharing PHI is the biggest challenge
- ▶ De-identification standards and practices vary widely
- ▶ Disparate interpretations of legal and regulatory requirements lead to widely different expectations and requirements
- ▶ Stakeholder input is vital
 - Developers
 - Researchers and other end users
 - Human subjects/patients
 - Security experts
 - Institutional Review Board (IRB) members
- ▶ Practices are evolving along with developments in federal guidance, emergence of standards/best practices, and novel issues
- ▶ International scientific collaborations increases complexity

Overview

- ▶ What is caBIG™?
- ▶ caBIG™ Structure and Oversight
- ▶ HIPAA-Related Issues
- ▶ **Issues Identified and Addressed**
- ▶ For More Information

Much confusion surrounds the interplay of HIPAA requirements and other Federal drivers.

	HIPAA Privacy Rule	Common Rule	FDA Rule	State Law
Brief Citation	45 CFR Parts 160 through 164, revised 8/14/02	45 CFR 46, revised 6/23/05	21 CFR Parts 50 and 56, revised 6/18/91	See National Conference of State Legislatures (genetic privacy) or Health Privacy Project
Applicability	Covered entities (health care providers, plans, and clearinghouses)	Persons and institutions receiving federal funds to conduct research	All research on human subjects conducted to seek a research or marketing permit	Varies
Permission mechanism	Protocol-specific authorization	Consent within a specified scope, including risks, benefits, rights, etc.	Consent within a specified scope with less defined content	Varies (e.g. notification)
Exceptions for privacy protections	"De-Identified" data	Thoroughly de-linked data? Or data not readily identifiable?	None	Varies
Oversight	Privacy Board	Institutional Review Board	Institutional Review Board	Varies

These federal requirements do not always provide explicit guidance; additionally, they are often interpreted as conflicting.

- ▶ Common Rule
 - Applies to all institutions conducting any federally-funded research on human subjects
 - August 2005 revision held that research on coded repository samples is not human subjects research; this shift requires interpretation by each institution
 - Boundaries of what research constitutes the “currently proposed project,” and therefore which research on samples is exempt
 - Allows broad patient consent

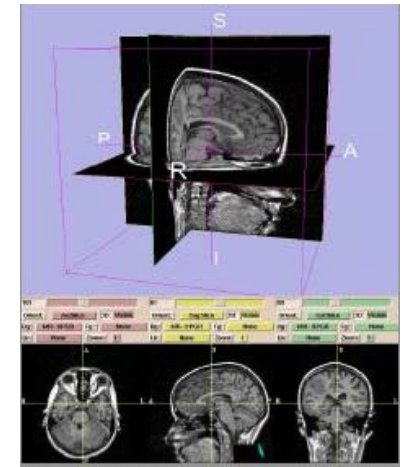
- ▶ HIPAA Privacy Rule
 - Only applies if protected health information (PHI) is transmitted by a covered entity
 - Other conditions that enable broader data sharing:
 - de-identified data
 - studies preparatory to research
 - studies involving decedents
 - “limited data set” use
 - Allows only protocol-specific patient authorization

Many groups are working toward greater efficiency and standardization.

- ▶ American Association of Medical Colleges (AAMC) seeks to [document the effects of HIPAA](#) on medical research, including the need to harmonize HIPAA with the Common Rule
- ▶ National Institutes of Health (NIH), through its [Clinical Research Policy Analysis and Coordination \(CRpac\) Program](#), is developing policies that harmonize the impact of HIPAA and the Common Rule on research conducted by NIH investigators.
- ▶ National Cancer Institute (NCI), through its Office of Biorepositories and Biospecimens Research (OBBR), has [proposed standard operating procedures for biorepositories generally](#), including a [sample consent form](#)
- ▶ [Public Responsibility in Research & Medicine \(PRIM&R\)](#) has been working on a document proposing harmonization of the Rules (originally proposed for January 2006); [other resources are available](#)
- ▶ International Society for Biological and Environmental Repositories (ISBER) [met to discuss these issues in Bethesda, MD in 2006](#): Action items have continued as private discussions
- ▶ FasterCures, a private organization dedicated to removing barriers to medical research, sponsors [BioBankCentral](#), a Web-based information source for researchers, advocates and the public.

DSIC WS has also responded to inquiries from other caBIG™ workspaces, such as the caBIG™ Imaging Workspace

- ▶ “A covered entity may determine that health information is not individually identifiable health information only if...[t]he following identifiers ... are removed: ... 18) Full face photographic images and other **comparable** images....”
 - HIPAA Privacy Rule, 45 CFR 164.514(b)(2)(i)(Q) (emphasis added).
- ▶ caBIG™ will develop tools to share PET/CT scans and other images
- ▶ Advice provided: Consult IRB or Privacy Board for an institutional determination of whether images to be shared could be used “alone or in combination with other information to identify an individual who is a subject of the information” and therefore constitute an identifier.
 - HIPAA Privacy Rule, 45 CFR 164.514(b)(2)(i)(Q)).



DSIC WS provided advice to members of caBIG™'s Tissue Banks and Pathology Tools Workspace (TBPT WS) as they integrated an automated de-identification tool.

- ▶ TBPT WS is developing caTIES, a tool that will assist with gene annotation
- ▶ caTIES includes a software component that can remove PHI from free text, thereby de-identifying it
- ▶ Automated de-identification is accomplished through the comparison of free-text reports with a dictionary of terms
- ▶ DSIC WS provided advice to participants on combining technical solutions for de-identification with non-technical solutions
 - Manual review
 - Training, education and awareness
 - Need-to-know/minimum necessary
 - Use of “honest broker” to re-identify records (and seek consent and authorization) for longevity studies, future recruitment, etc.

DSIC WS is developing security policies and procedures that include collecting specific boundary requirements from end users.

- ▶ Security and privacy needs are diverse within the caBIG™ community
- ▶ Tools vary in:
 - Complexity
 - Maturity of Information Model
 - Security/privacy parameters
 - Regulatory environments
 - Supporting technology requirements
- ▶ To clarify what was needed to support these domains, in Summer of 2005, caBIG™ commissioned a Security White Paper
 - Prepared by Booz Allen Hamilton with extensive input and review of the Architecture Workspace and NCICB
- ▶ Scope of paper was largely technical evaluation, with some comments on policy and administration issues
- ▶ Draft circulated in October 2005 with final version available in February 2006.

In response to the Security Technology Evaluation White Paper, a team was tasked with addressing its unmet recommendations.

Recommendations from the White Paper included:

- ▶ Develop business-oriented security use & abuse cases
 - Need input from IRBs, Compliance Officers, Honest Brokers, CIOs and other institutional executives, Bioethicists, etc.
- ▶ Vet the notion of employing Federated Identity Management
- ▶ Develop caBIG™ governance policies
 - Success involves multiple layers (i.e., trust, identity vetting, guidelines, data standards, firewalls, physical security, etc.)
- ▶ Involve multiple workspaces and stakeholders in policy development
- ▶ Identify the minimum security requirements from regulatory mandates
- ▶ Develop a Proof-of-Concept implementation
- ▶ Consider the maturity of technologies
- ▶ Consider separating regulated and non-regulated environments

Some standards of the HIPAA Security Rule may require coordination among caBIG participants.

- ▶ Some HIPAA requirements may dictate needed functionalities of tools or infrastructure: The Security White Paper supported this idea
- ▶ Other requirements may require coordinated administrative activities such as:
 - A recommended *governance structure* for security
 - Recommended processes for *risk assessment and management*
 - A process enabling *information system activity review*
 - A process for *on going policy and operations review* involving end users and stakeholders
 - Requirements for *external audit review & associated policies*
 - A process for managing *security incidents & events*
 - A process for *management, review, and modification of interconnection security agreements*

The real challenges of cross-institutional data sharing are political and cultural, not technical.

- ▶ Sharing information among institutions will require the approval of institutional review boards (IRBs)
- ▶ An agreement on minimum security acceptable to a broad range of IRBs and other compliance officials will be critical to receiving approval (“interconnection security agreement”)
- ▶ Healthcare security problems are complex as epitomized by:
 - **Infrastructure Gaps:** Some institutions are sharing data via e-mail attachments while others have more sophisticated biomedical informatics systems
 - **Scientific vs. Engineering Mindset:** Enthusiastic about technologies; needs to understand the importance of having an integrated engineering process
 - **Regulatory Compliance:** IRBs tend to be both conservative and disparate in interpretation of federal requirements.

caBIG™ Security Program Goals

- ▶ Major goal is to develop a framework for security engineering for the caBIG™ project as a whole
- ▶ Targets Cancer Centers which are the initial four adopters of caTIES
 - Washington University, U. Pittsburgh Medical Center, Thomas Jefferson, U Penn
- ▶ Focus on involving regulatory and other “business users” at the Cancer Centers
 - IRB members
 - Compliance officers
- ▶ Deliverables:
 - Capstone governance structure framework and documents
 - Security refinement processes
 - Interconnection security agreement among adopters
 - Policy and procedures sufficient to operate caTIES at individual Cancer Centers
- ▶ Cross-cutting joint effort between Architecture, VCDE, TBPT, DSIC Workspaces

Policies and procedures are being developed through a process that addresses specific needs of stakeholders.

- ▶ A community of stakeholders and subject matter experts assembled at a face-to-face meeting in June, 2006
- ▶ Participants included experts in the fields of security, law, IRB operations, compliance officers, bioinformaticists, and others
- ▶ After soliciting input from all attendees, the team developed scenarios to be used as a framework for open-ended discussions with stakeholders
- ▶ The scenarios will be used to identify security requirements *boundary conditions*
- ▶ All deliverables (survey-like instruments, ultimately drafted policies etc.) will be:
 - Subject to community review
 - Used to inform the development of the technical infrastructure

Current Status

- ▶ The outcome of the process was the development of four generic scenarios to be used as a framework for open-ended interviews of:
 - IRB staff
 - Compliance officers
 - Security officers
 - IT staff
 - Others as appropriate
- ▶ Four major topics identified
 - Locus of Control
 - Auditing
 - Consenting
 - De-identification



Data Collection Plan for Security Policy and Process Interviews

- ▶ Task team has identified relevant stakeholders at the institutions included in the study
- ▶ Set up interviews with stakeholders, using usage scenarios and related questions to further define requirements
 - Scenarios will be sent ahead of time to interviewees
 - Also sent in advance: A glossary of terms related to privacy and security
- ▶ Each interview will start with 10 minute presentation on caBIG™ and caTIES
- ▶ Summaries of interviews will be sent back to the interviewee to review and approve
 - Review may be accompanied by specific requests for clarification if questions arise during analysis
- ▶ Data will be summarized and analyzed by institution and across institutions.

Overview

- ▶ What is caBIG™?
- ▶ caBIG™ Structure and Oversight
- ▶ HIPAA-Related Issues
- ▶ Issues Identified and Addressed
- ▶ [For More Information](#)

Your involvement in caBIG™ is welcomed!

- ▶ Data Sharing and Intellectual Capital Workspace meets 2-3 PM EST:
 - First Thursday of each month
 - Second and fourth Thursday of each month (Proprietary/Intellectual Property Issues)
 - Second and fourth Monday of each month (Regulatory issues, e.g. HIPAA)

- ▶ Contacts:

Working Group Facilitator (NCI)

Wendy Patterson, J.D.

(301) 435-3110

pattersw@mail.nih.gov

Working Group Coordinator (BAH)

Dan Steinberg, J.D.

(703) 377-1261

steinberg_daniel@bah.com

More on caBIG™ at:

<http://caBIG.nci.nih.gov/>

Save the Date for caBIG™'s Annual Meeting 2007!

- ▶ February 5 - 7, 2007
- ▶ Marriott Wardman Park, Washington, DC
- ▶ Sessions include
 - Plenaries
 - Dozens of break outs
 - Dozens of demonstrations, posters, and exhibits
 - Hands-on introduction to caBIG™ tools (the popular “caBIG™ Hackathon”)
- ▶ Tailored sessions for newcomers February 5 and throughout the conference
- ▶ <https://cabig.nci.nih.gov/2007caBIGconference/>

| Questions/Discussion?

| Supplemental Materials

A Sample of the caBIG™ Participant Community

9Star Research
Albert Einstein
Ardais
Argonne National Laboratory
Burnham Institute
California Institute of Technology-JPL
City of Hope
Clinical Trial Information Service (CTIS)
Cold Spring Harbor
Columbia University-Herbert Irving
Consumer Advocates in Research
and Related Activities (CARRA)
Dartmouth-Norris Cotton
Data Works Development
Department of Veterans Affairs
Drexel University
Duke University
EMMES Corporation
First Genetic Trust
Food and Drug Administration
Fox Chase
Fred Hutchinson
GE Global Research Center
Georgetown University-Lombardi
IBM
Indiana University
Internet 2
Jackson Laboratory
Johns Hopkins-Sidney Kimmel
Lawrence Berkeley National Laboratory
Massachusetts Institute of Technology
Mayo Clinic
Memorial Sloan Kettering
Meyer L. Prentis-Karmanos
New York University

Ohio State University-Arthur G. James/Richard Solove
Oregon Health and Science University
Roswell Park Cancer Institute
St Jude Children's Research Hospital
Thomas Jefferson University-Kimmel
Translational Genomics Research Institute
Tulane University School of Medicine
University of Alabama at Birmingham
University of Arizona
University of California Irvine-Chao Family
University of California, San Francisco
University of California-Davis
University of Chicago
University of Colorado
University of Hawaii
University of Iowa-Holden
University of Michigan
University of Minnesota
University of Nebraska
University of North Carolina-Lineberger
University of Pennsylvania-Abramson
University of Pittsburgh
University of South Florida-H. Lee Moffitt
University of Southern California-Norris
University of Vermont
University of Wisconsin
Vanderbilt University-Ingram
Velos
Virginia Commonwealth University-Massey
Virginia Tech
Wake Forest University
Washington University-Siteman
Wistar
Yale University
Northwestern University-Robert H. Lurie