



**caBIG™** cancer Biomedical  
Informatics Grid™

# caBIG™, caTissue, and Achieving Silver-Level Compatibility

***Ian Fore D.Phil.***

*Associate Director, Biorepository and  
Pathology Informatics  
NCI Center for Biomedical Informatics  
and Information Technology*

January 28, 2008

# Unprecedented Potential for Progress

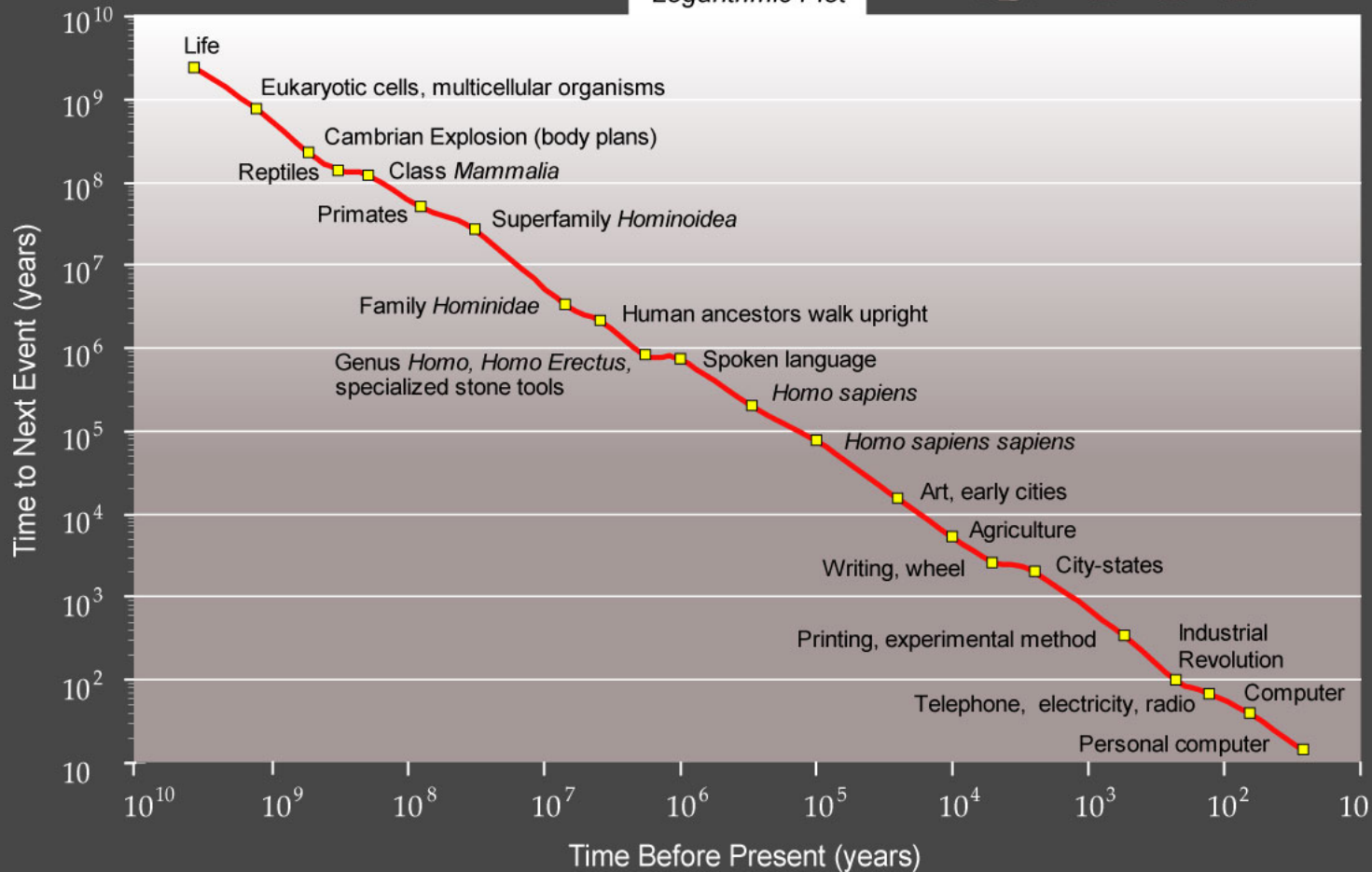
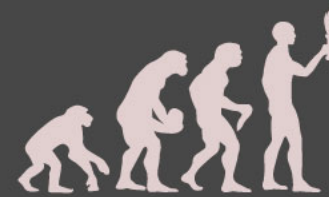


- Technological change is exponential, not linear
  - “We won’t experience 100 years of progress in the 21st century – it will be more like 20,000 years of progress (at today’s rate).”
  - Ray Kurzweil, *The Law of Accelerating Returns*
- Scientific knowledge will double in the next 3 years
- Biologic knowledge will double in the next 5 years
- The sum of all human knowledge is just 1% of what it will be in the year 2050

# Kurzweil - *The Singularity is Near*

## Countdown to Singularity

Logarithmic Plot



# Outline



- Informatics in the Best Practices
  - Functions
  - Development
  - Security
- caBIG
  - Why caBIG compatibility?
  - What does it take?
  - Support programs
- caBIG and Tissue Banks
  - caTissue Core
  - caTIES
  - caTissue Suite

# Informatics and the Best Practices

# Functionality - within the repository



- Aim to identify the major areas that biorepository software should support
  - Participant registration
  - Consent tracking
  - Biospecimen collection
  - Processing
  - Storage
  - Distribution
  - Specimen QA & QC
  - Security
    - Who can see/access/order which specimens
  - Reporting
    - Help biorepository managers to manage their collection...
    - What's past its "use by" date

# Specimen tracking



- Treat each physically distinct entity as a different specimen
  - ... with its own identifier
- Maintain parent-child relationship of specimens
  - Aliquoting
  - Molecular extracts
- Linkage to physical labelling
  - Barcodes

# Integration



- Seek to integrate with those clinical data systems that provide useful clinical annotation of stored biospecimens
  - Pathology systems
    - Anatomic pathology
    - Clinical pathology
  - Cancer registries
  - Operating room systems
    - Anesthesia records



- To match the special nature of human biospecimens systems should be protected by adequate security
  - Physical access to systems
  - Back up of systems containing often irreplaceable resources
  - Login protections
  - Role based security that only allows access to authorized information
- National Institute of Standards and Technology (NIST)
- *Risk Management Guide for Information Technology Systems*
  - Use to help determine level of risk for a particular system
  - Set security mechanisms to match the risk

# Regulatory and sharing requirements



- Health Insurance Portability and Accountability Act (HIPAA)
- Human Subjects Research - CFR Title 45 Part 46
- FDA requirements - CFR Title 21 Part 11
- NIH Principles and Guidelines for Sharing of Biomedical Resources
- NIH Data Sharing Policy

# Choosing a system



- **Use of structured information**
  - **Databases instead of free text**
- **Build vs Buy**
  - Understand the true costs
  - Plan for the future
    - Establishing a resource that will realize its value in years
    - Make sure the informatics will still be around then
  - Open source

# Some Biospecimen Software Tools



Name	Organization	License	caBIG Compatibility
<a href="#">Sapphire R4</a>	LabVantage	Commercial	Bronze - under review
<a href="#">Freezerworks</a>	Dataworks Development	Commercial	Bronze
<a href="#">LabMatrix</a>	BioFortis	Commercial	Unknown
<a href="#">caTissue Core</a>	caBIG	Open source	Silver
<a href="#">Waban LIMS</a>	Waban	Commercial	Unknown
<b>Biological Specimen Inventory System (BSI)</b>	IMS	Commercial	Bronze
<a href="#">BIGR - Biomaterials and Information for Genomic Research</a>	GulfStream Bioinformatics	Commercial	Unknown
<b>Labrador</b>	NCI	Unknown	Unknown
<a href="#">Biomaterial Tracking and Management - Research</a>	Daedalus Software	Commercial	Unknown
<a href="#">TissueMetrix</a>	Artificial Intelligence in Medicine Inc.	Commercial	Unknown
<a href="#">Oncore</a>	PercipEnz	Commercial	Unknown

# If you're going to build...



- End user involvement
- Use cases
- Follow a system development methodology
  - E.g. Unified Process
- Understand the true costs
  - *Brooks Law*
  - *"Adding manpower to a late software project makes it later."*
  - Costs of maintenance and support
  - The Mythical Man Month - Fred P Brooks, Jr.
- Strive for Capability Maturity Model Level 3



**caBIG**<sup>™</sup> cancer Biomedical  
Informatics Grid<sup>™</sup>

# caBIG<sup>™</sup> – NCI's Approach to the IT Infrastructure



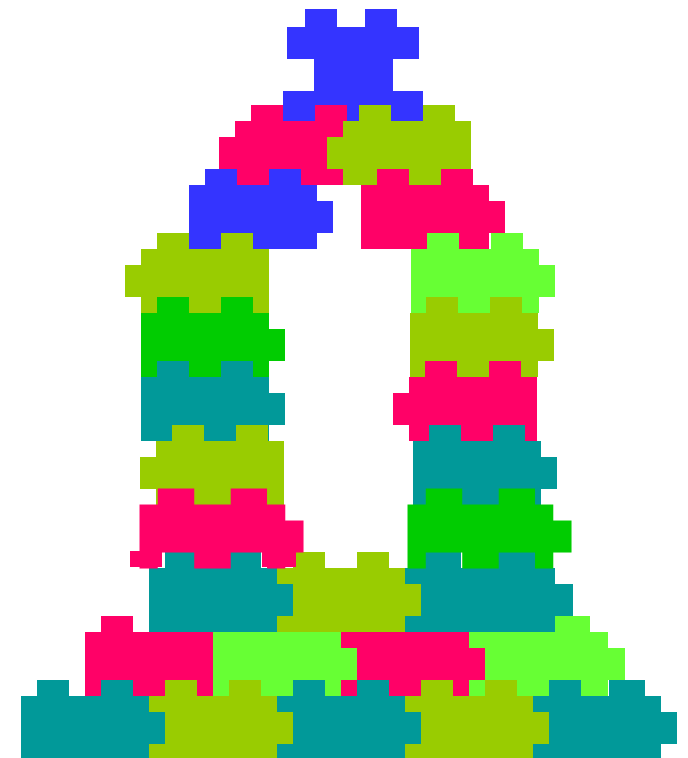
**NATIONAL  
CANCER  
INSTITUTE**

<b>Clinical Trial Management Systems</b>	Addresses the need for consistent, open and comprehensive tools for clinical trials management.
<b>Integrative Cancer Research</b>	Provides tools and systems to enable integration and sharing of information.
<b>Tissue Banks &amp; Pathology Tools</b>	Provides for the integration, development, and implementation of tissue and pathology tools.
<b><i>In vivo</i> Imaging</b>	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.	<b>Vocabularies &amp; Common Data Elements</b>
Developing architectural standards and architecture necessary for other workspaces.	<b>Architecture</b>

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

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



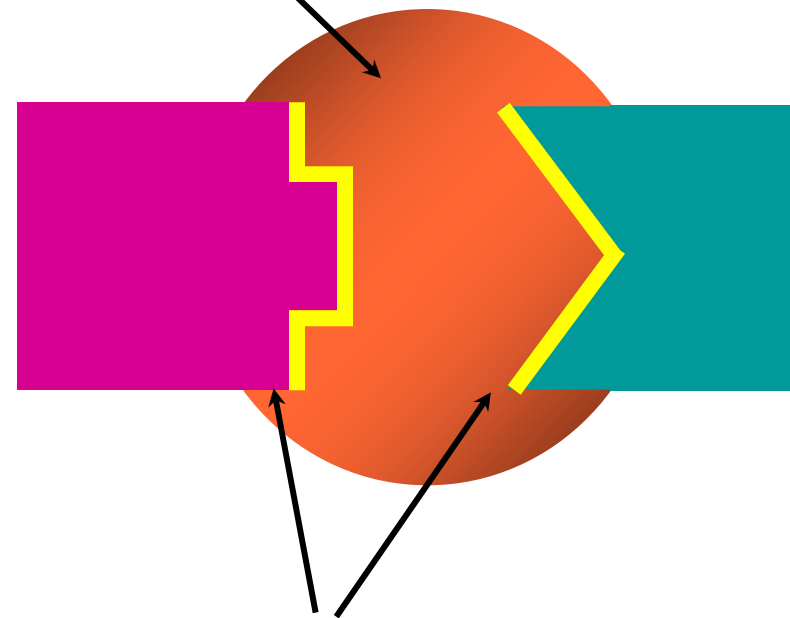


# Boundaries and Interfaces



- **focus on boundaries of how how things fit together, not on the internal details**
- **once they're built: assume that will be diverse & changing**

The glue that binds parts together is middleware infrastructure



Shape of boundary is defined in APIs



# Interoperability benefits for Biospecimens



- Virtual biorepositories
  - The ability to search networked biospecimen databases
- Support for multi site studies
  - Specimens collected at multiple locations
  - E.g. Prostate SPORE Biomarker study
- Rare diseases
  - Statistical power only reached through pooling resources
- Biospecimens are the raw material for molecular analysis platforms
  - Need to exchange data with their informatics systems
- Linkage to clinical data on biospecimens
  - Key data on specimens and patients exists in other systems
- Development of common practices for biospecimen handling
  - Establishing common data elements will support this

# Biospecimen Best Practices Toolkit: Written Background Materials



- Items for distribution at meetings and through the OBBR Web site
  - ✓ NCI Best Practices for Biospecimen Resources
  - ✓ Biospecimen Basics: An Overview of the NCI Best Practices for Biospecimen Resources
  - ✓ **Implementing caBIG™ for Biospecimen Resources: An Overview**
  - ✓ **Implementing caBIG™ for Biospecimen Resources: Next Steps**
  - ✓ Providing Your Tissue for Research
  - ✓ Other biospecimen-related articles, publications, and news stories of interest

# Options for Biospecimen Resources



Option	Operating Scenario	Recommended Solution
1	Your resource has a paper-based system or a homegrown tool that would not be painful to abandon.	Adopt caTissue Core, and migrate existing electronic data to new tool.
2	Your resource has an existing basic tool that you want to keep. Examples: Access or mySQL database.	Become caBIG™ compatible by installing caTissue Core, and then mapping your tool to it. Connect from your tool to the outside world through caTissue Core.
3	Your resource has an existing informatics tool - more complex than a simple database, with separate modules for reporting and storage (e.g., standard reports are a small extract from a larger database system).	Make the existing tool caBIG™ compatible for your standard reports only. Data generated in required reports would be caBIG™ compatible; the underlying data need not be.
4	Your resource has an existing complex informatics tool (like Option 3), but your reporting needs vary greatly, and you would like to have the entire system caBIG™ compatible for maximum flexibility.	Make the full database compatible, by creating an interface that maps the existing tool's data structures to caBIG™ standards. This is the highest investment solution.

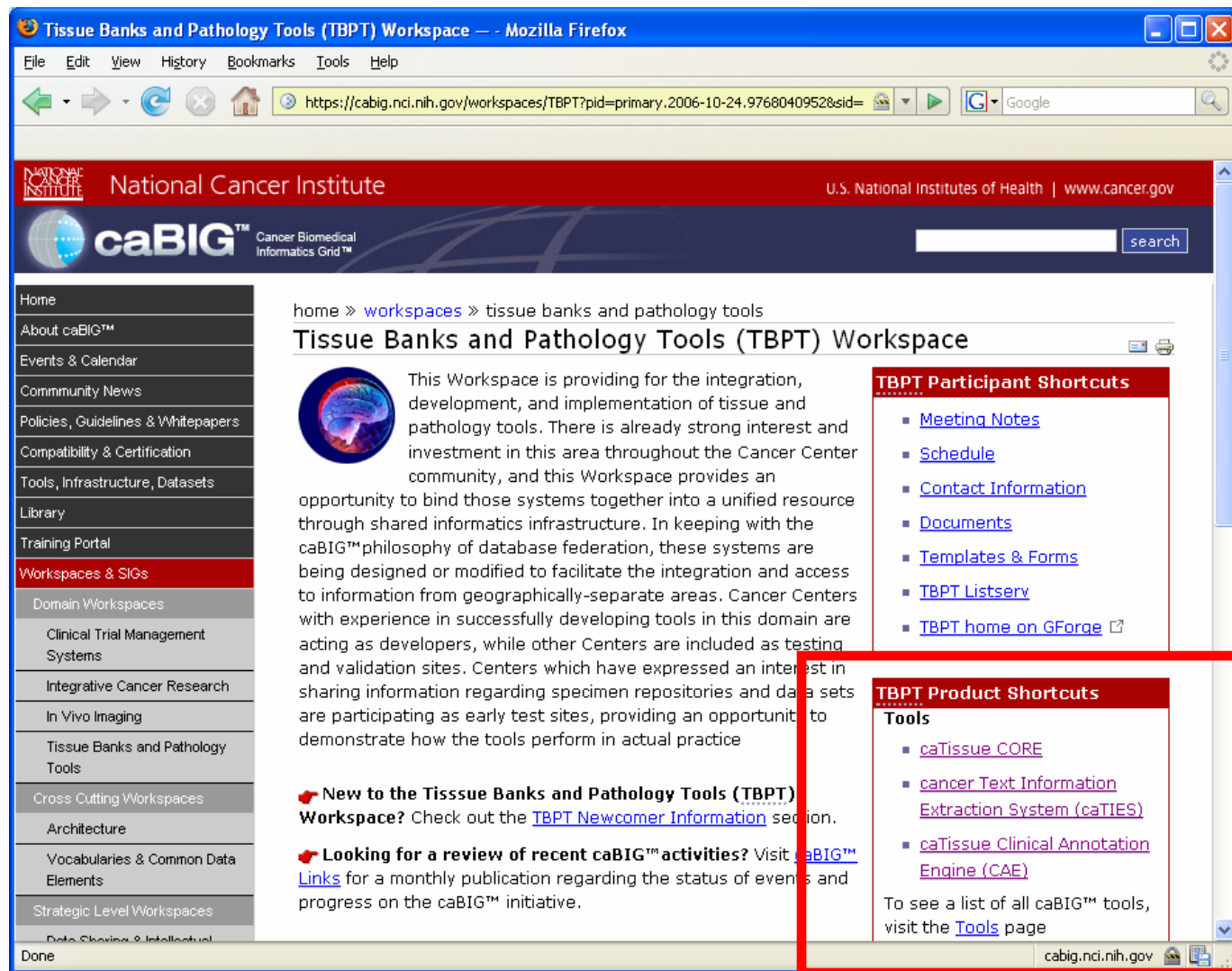
# caBIG Biorepository and Pathology Tools

# caBIG™ Benefits Biospecimen Resource Leaders, Researchers and Advocates



- **Leveraging existing software tools available through caBIG™ reduces software development costs for biospecimen resources**
- **The shared standards governing caBIG™ allows even small repositories to advertise their presence and specimen/data availability – AND – learn what others have to offer**
- **Your own research comes first – you can select the data to share**
- **caBIG™ tools already have built-in security and privacy considerations – enhancing patient confidence**
- **Increased data sharing increases the effectiveness and efficiency of cancer research – helping individual scientists, the cancer research community, and ultimately the cancer patient**

**The willingness of cancer patients to share tissue is fundamental to cancer research. Our willingness to share biospecimen-related data is critical to maintaining public trust.**



Tissue Banks and Pathology Tools (TBPT) Workspace — - Mozilla Firefox

File Edit View History Bookmarks Tools Help

https://cabig.nci.nih.gov/workspaces/TBPT?pid=primary.2006-10-24.9768040952&sid=...

National Cancer Institute U.S. National Institutes of Health | www.cancer.gov

caBIG™ Cancer Biomedical Informatics Grid™

Home

About caBIG™

Events & Calendar

Community News

Policies, Guidelines & Whitepapers

Compatibility & Certification

Tools, Infrastructure, Datasets

Library

Training Portal

**Workspaces & SIGs**

Domain Workspaces

Clinical Trial Management Systems

Integrative Cancer Research

In Vivo Imaging

Tissue Banks and Pathology Tools

Cross Cutting Workspaces

Architecture

Vocabularies & Common Data Elements


Strategic Level Workspaces

Data Sharing & Intellectual

Done

home » [workspaces](#) » tissue banks and pathology tools

## Tissue Banks and Pathology Tools (TBPT) Workspace

 This Workspace is providing for the integration, development, and implementation of tissue and pathology tools. There is already strong interest and investment in this area throughout the Cancer Center community, and this Workspace provides an opportunity to bind those systems together into a unified resource through shared informatics infrastructure. In keeping with the caBIG™ philosophy of database federation, these systems are being designed or modified to facilitate the integration and access to information from geographically-separate areas. Cancer Centers with experience in successfully developing tools in this domain are acting as developers, while other Centers are included as testing and validation sites. Centers which have expressed an interest in sharing information regarding specimen repositories and data sets are participating as early test sites, providing an opportunity to demonstrate how the tools perform in actual practice

**New to the Tissue Banks and Pathology Tools (TBPT) Workspace?** Check out the [TBPT Newcomer Information](#) section.

**Looking for a review of recent caBIG™ activities?** Visit [caBIG™ Links](#) for a monthly publication regarding the status of events and progress on the caBIG™ initiative.

### TBPT Participant Shortcuts

- [Meeting Notes](#)
- [Schedule](#)
- [Contact Information](#)
- [Documents](#)
- [Templates & Forms](#)
- [TBPT Listserv](#)
- [TBPT home on GForge](#)

### TBPT Product Shortcuts

#### Tools

- [caTissue CORE](#)
- [cancer Text Information Extraction System \(caTIES\)](#)
- [caTissue Clinical Annotation Engine \(CAE\)](#)

To see a list of all caBIG™ tools, visit the [Tools](#) page

cabig.nci.nih.gov

caTissue Core — Mozilla Firefox

File Edit View History Bookmarks Tools Help

https://cabig.nci.nih.gov/tools/catissuecore

National Cancer Institute U.S. National Institutes of Health | www.cancer.gov

caBIG™ Cancer Biomedical Informatics Grid™ search

Home

About caBIG™

Events & Calendar

Community News

Policies, Guidelines & Whitepapers

Compatibility & Certification

Tools, Infrastructure, Datasets

Library

Training Portal

**Workspaces & SIGs**

Domain Workspaces

Clinical Trial Management Systems

Integrative Cancer Research

In Vivo Imaging

Tissue Banks and Pathology Tools

Done

home » [tools](#) » [catissue core](#)

## caTissue Core

The screenshot shows the caTissue Core web application interface. The main content area displays a 'NEW DISTRIBUTION PROTOCOL' form with the following fields:

- Principal Investigator:** Wilson, Mark
- Title:** Super Prostate Microarray Experiment
- Short Title:** Prostate Microarray
- PIB ID:** 05-7777
- Start Date:** (calendar icon)
- End Date:** (calendar icon)
- Number of Participants Anticipated:** 250
- Description URL:** (text input)

Below the form is a 'SPECIMEN REQUIREMENTS' table:

#	Class	Type	Tissue Site	Pathology Status	Quantity (est)
1	Tissue	Slide	PROSTATE GLAND	Intact	25
2	Molecule	RNA	PROSTATE GLAND	Non-Intact	01 mg
3	Molecule	RNA	PROSTATE GLAND	Intact	01 mg

Quick Links to Page Contents

- [Product Summary](#)
- [Demo Files & Exercises](#)
- [Documentation & Training](#)
- [Tool Download Files](#)
- [End User Support Resources](#)
- [Related caBIG™ Tools](#)
- [Developer Institution\(s\)](#)
- [Adopter Institution\(s\)](#)
- [Related Articles](#)

cabig.nci.nih.gov



# cancer Text Information Extraction System



**cancer Text Information Extraction System (caTIES) — - Mozilla Firefox**

File Edit View History Bookmarks Tools Help

https://cabig.nci.nih.gov/tools/caaties

NATIONAL CANCER INSTITUTE National Cancer Institute U.S. National Institutes of Health | www.cancer.gov

caBIG™ Cancer Biomedical Informatics Grid™ search

Home

About caBIG™

Events & Calendar

Community News

Policies, Guidelines & Whitepapers

Compatibility & Certification

Tools, Infrastructure, Datasets

Library

Training Portal

**Workspaces & SIGs**

Domain Workspaces

Clinical Trial Management Systems

Integrative Cancer Research

In Vivo Imaging

Tissue Banks and Pathology Tools

Cross Cutting Workspaces

Done

home » [tools](#) » cancer text information extraction system (caaties)

## cancer Text Information Extraction System (caTIES)

The screenshot shows the caTIES web application interface. On the left, there is a navigation menu with various options. The main content area displays a search workflow diagram with steps like 'Start Discussion', 'Sex', 'Age', and 'Search Terms'. To the right, there is a list of patient records with columns for Patient ID, Name, and other details. Below the list, there is a section for 'Available Tools' and a 'Quick Links to Page Contents' sidebar with links to Product Summary, Demo Files & Exercises, Documentation & Training, Tool Download Files, End User Support Resources, Related caBIG™ Tools, Developer Institution(s), Adopter Institution(s), and Related Articles.

Quick Links to Page Contents

- [Product Summary](#)
- [Demo Files & Exercises](#)
- [Documentation & Training](#)
- [Tool Download Files](#)
- [End User Support Resources](#)
- [Related caBIG™ Tools](#)
- [Developer Institution\(s\)](#)
- [Adopter Institution\(s\)](#)
- [Related Articles](#)

cabig.nci.nih.gov

National Cancer Institute



**caBIG**<sup>™</sup> cancer Biomedical  
Informatics Grid<sup>™</sup>

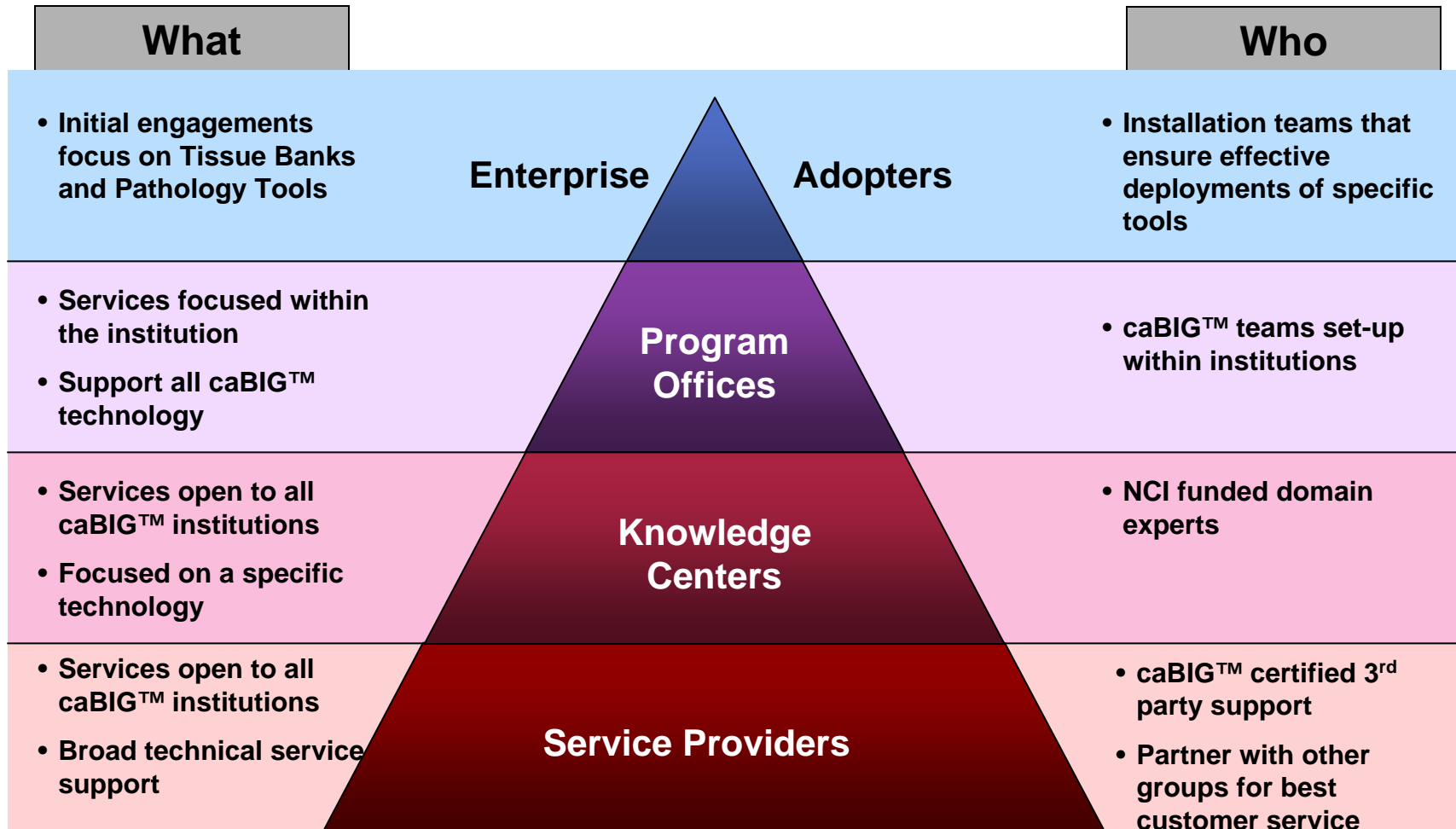
# Enterprise Support Network

U.S. DEPARTMENT  
OF HEALTH AND  
HUMAN SERVICES

National Institutes  
of Health

December 2006

# Facilitating Next Generation Adoption



Ongoing Tool Development, Adoption and Participation

# Stay Connected to the caBIG™ Community



For more background  
visit:

[caBIG.cancer.gov](http://caBIG.cancer.gov)

To join the caBIG™  
technical effort visit:

[caBIG.nci.nih.gov](http://caBIG.nci.nih.gov)

The screenshot shows the caBIG website homepage. At the top, there is a red header with the National Cancer Institute logo and the text "National Cancer Institute" and "U.S. National Institutes of Health | www.cancer.gov". Below this is a dark blue navigation bar with the caBIG logo and a search box. The main content area features a large banner with the text "caBIG™: power of connection™" and an illustration of a globe with scientists working at computers. Below the banner is a section titled "How can information technology help fight cancer?" with a paragraph of text and a "Read More" link. To the right, there are three smaller sections: "caBIG In Action" with a video thumbnail and a "Watch Video" link, "News Center" with a paragraph of text and a "Read More" link, and "Already a caBIG participant?" with a paragraph of text and a "Click Here" link. At the bottom, there is a footer with navigation links: "Home | Text-Only Version | Contact Us | Policies | Accessibility | Search | Site Map", the text "A Service of the National Cancer Institute", and logos for the National Cancer Institute, the Department of Health and Human Services, and FIRSTGOV.