



caBIG[™] cancer Biomedical
Informatics Grid[™]

Tools for Tracking and Accessing High-Quality Biospecimens caBIG[™], caTissue, and Achieving Silver-Level Compatibility

Ken Buetow, Ph.D.
*NCI Associate Director
Bioinformatics and
Information Technology*

June 18, 2007

- 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

- 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

- 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

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

- 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
- 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

- 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

- **Use of structured information**
- **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
- **If you're going to build...**
 - End user involvement
 - Use cases
 - Follow a system development methodology
 - E.g. Unified Process
 - Strive for CMMI Level 3



caBIG[™] cancer Biomedical
Informatics Grid[™]

caBIG[™] – NCI's Approach to the IT Infrastructure



**NATIONAL
CANCER
INSTITUTE**

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.
<i>In vivo</i> 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 tutorials, workshops, training programs.

Strategic Planning

Assists in identifying strategic priorities for the development and evolution of the caBIG™ effort.

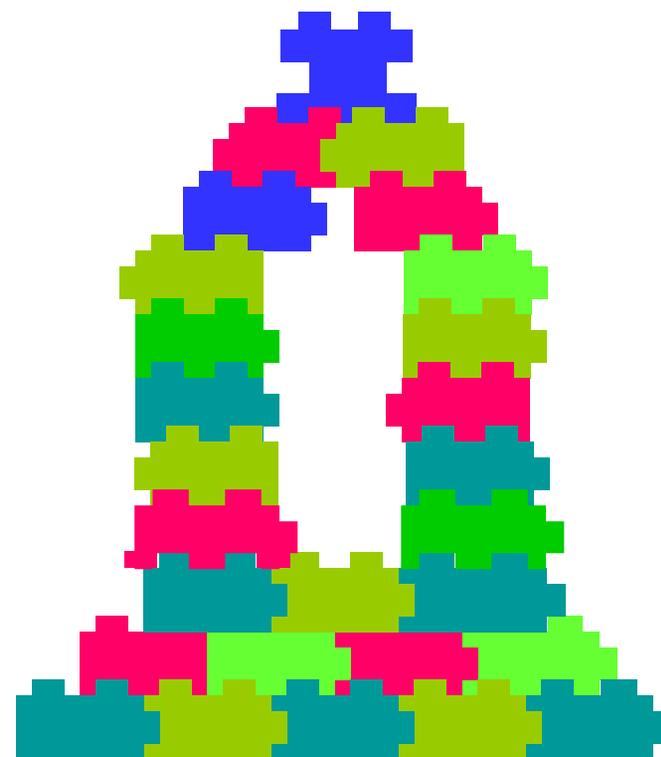
Cancer Research Landscape



- Integrated Systems
 - Homegrown/Commercial
 - Smooth navigation between applications
 - Difficult to expand/extend
 - Large IT staff
 - \$10M's invested
- Heterogeneous Systems
 - Complex mix of commercial and homegrown components (may be composed of dozens of components)
 - No common interfaces
 - Medium size IT staff
 - \$1M's invested
- Informal/ no systems
 - Use of productivity applications (e.g. Excel, Access)
 - Complex manual processes
 - Small or no IT staff
 - \$100K's invested

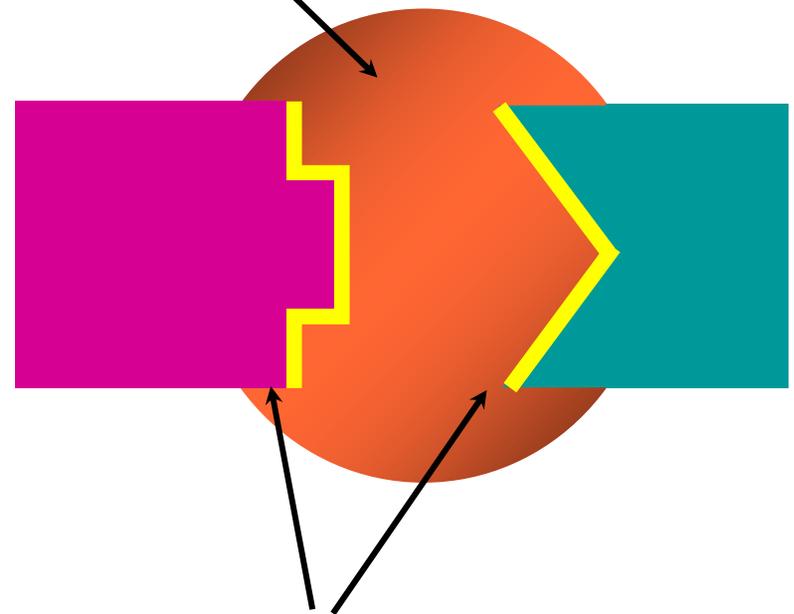


- **Modules that address specific needs**
- **Connect through defined Electronic interfaces**
- **Use of international data standards**



- **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



- 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

Implementing caBIG™ for Biospecimen Resources: An Overview



- One-page document with broad overview of caBIG™ for Biospecimen Resources
- Target audiences include:
 - ✓ Patients and Patient Advocates
 - ✓ NCI-designated Cancer Center Directors
 - ✓ Strategic Thinkers at Institutes and Biospecimen Resources
- Topics Covered:
 - ✓ What is caBIG™ and What does it offer Biospecimen Resources?
 - ✓ Strategic Considerations
 - ✓ Benefits to patients and advocates
 - ✓ Benefits to researchers and resource directors

Implementing caBIG™ for Biospecimen Resources: Next Steps



- **9-page document highlighting “The Road to caBIG™ Compatibility”**
- **Designed for resources interested in the specific steps required for implementing caBIG™ compatibility**
- **Addresses public comments and frequently asked questions**

- **Setting the Stage:**
 - ✓ Core Concepts
 - ✓ How does caBIG™ compatibility work
 - ✓ FAQs

- **Turning to Solutions:**
 - ✓ Available software tools
 - ✓ Overview of Alternatives
 - ✓ Skills, Technology, and Resources Required

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™ 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
Architecture
Vocabularies & Common Data Elements
Strategic Level Workspaces
Data Sharing & Intellectual Property

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

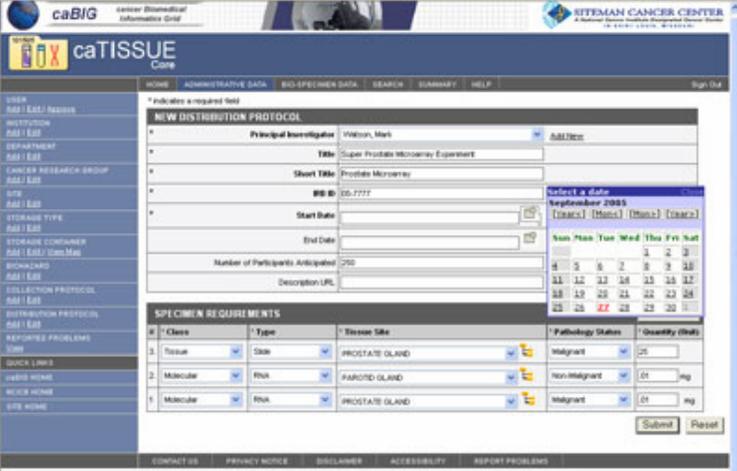
In Vivo Imaging

Tissue Banks and Pathology Tools

Done

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

caTissue Core



The screenshot shows the 'NEW DISTRIBUTION PROTOCOL' form in the caTissue Core application. The form includes fields for Principal Investigator (Wilson, Mark), Title (Super Prostate Microarray Experiment), Short Title (Prostate Microarray), and a calendar for selecting a start date (September 2005). Below the form is a table for 'SPECIMEN REQUIREMENTS' with columns for Class, Type, Tissue Site, Pathology Status, and Quantity (ml/g).

#	Class	Type	Tissue Site	Pathology Status	Quantity (ml/g)
1	Tissue	Core	PROSTATE GLAND	Intact	25
2	Molecular	RNA	PAROTID GLAND	Non-Intact	01 mg
3	Molecular	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 (caTIES) — - Mozilla Firefox

File Edit View History Bookmarks Tools Help

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

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

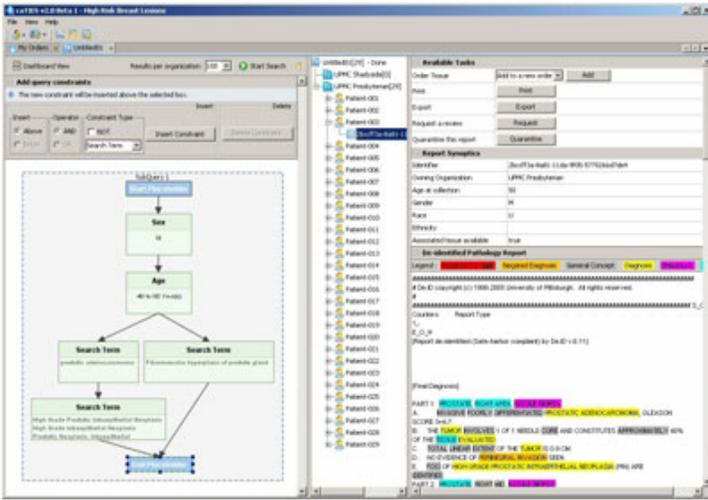
Tissue Banks and Pathology Tools

Cross-Cutting Workspaces

Done

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

cancer Text Information Extraction System (caTIES)



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

caTissue Clinical Annotation Engine



caTissue Clinical Annotation Engine (CAE) — - Mozilla Firefox

File Edit View History Bookmarks Tools Help

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

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](#) » [cattissue clinical annotation engine \(cae\)](#)

caTissue Clinical Annotation Engine (CAE)

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

What's Required: Skills and Technology Needed



- Some specialized IT skills are required to adopt caBIG™ tools, and to make an existing tool caBIG™ compatible.
 - ✓ Required skills: Intermediate Java and/or .NET development skills, with experience installing dependent software sets (e.g., Java, JBOSS, MySQL), administering systems, configuring line code, and similar tasks of moderate technical complexity.
 - ✓ Options 1, 2 or 3 described above will take a few weeks – or less.
 - ✓ Full system conversion (Option 4) is likely to take more investment – on the order of months rather than weeks.
 - ✓ We don't have precise cost estimates for recommended options at this time.

You do not have to hire full-time staff, or invest in an IT lab, to fulfill bioinformatics best practices.

You may need to “borrow” personnel at your institution with specialized IT skills for initial installation and routine maintenance.

The technical environment will include application and database servers; likely to be available at your institution without heavy investment by your lab if you don't already have them.



caBIG[™] cancer Biomedical
Informatics Grid[™]

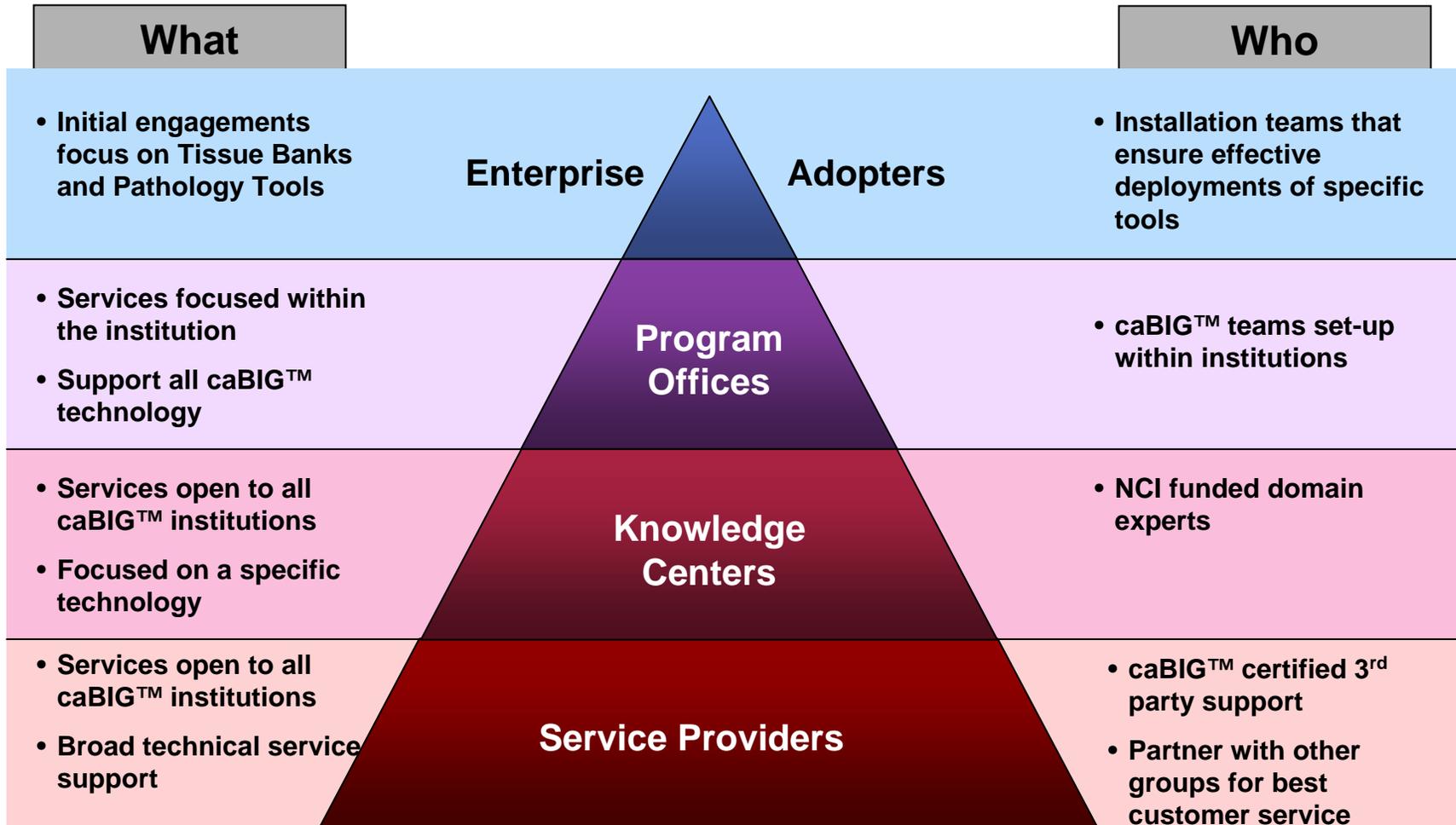
Enterprise Support Network

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

To join the caBIG™
technical effort visit:

caBIG.nci.nih.gov

The screenshot shows the National Cancer Institute's caBIG website. The header includes the NCI logo, the text "National Cancer Institute", and "U.S. National Institutes of Health | www.cancer.gov". The caBIG logo is prominently displayed. A search bar is located in the top right. Below the header is a navigation menu with four items: "Cancer and Molecular Medicine", "caBIG: A New Network", "caBIG and You", and "Media Center". The main content area features a large graphic with the text "caBIG™: power of connection™" and an illustration of a globe with scientists working at computers. Below this graphic is a section titled "How can information technology help fight cancer?" with a "replay" icon. To the right, there are three sidebar sections: "caBIG In Action" with a video thumbnail and a "Watch Video" link, "News Center" with a paragraph of placeholder text and a "Read More" link, and "Already a caBIG participant?" with a "Click Here" link. The footer contains 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.