



caBIG Program Update

In This Issue >

Meet the Domain Workspaces

Tissue Banks & Pathology Tools Workspace

TBPT Participants and Meetings

Current Activities of the TBPT Workspace

Next Steps

How to get involved with the TBPT Workspace

caBIG Information and Resources

Raising Questions or Concerns

Contact

Mark Adams, PhD

adamsm@mail.nih.gov

MESSAGE FROM NCICB

Welcome to the September edition of the caBIG Program Update. NCI is excited and pleased at the continued progress and success we've seen with the caBIG initiative. To date, we have a total of 39 signed contracts and verbal agreements from an additional 4 Cancer Centers to sign in the coming days. The caBIG Contracts Staff has issued over 27 task orders against the signed contracts and has an additional 9 task orders in the pipeline.

As activities amongst the Workspaces and Working Groups continue to gain momentum, initial face-to-face meetings are being held. The Clinical Trial Management Systems, Architecture, Vocabularies and Common Data Elements, and Integrative Cancer Research Workspaces held their first face-to-face meetings this summer, and the Tissue Banks and Pathology Tools Workspace held a developer summit in September. Two additional face-to-face meetings are planned for the coming months - a joint meeting between the Architecture and Vocabularies and Common Data Elements Workspaces, and a meeting for the Strategic Planning Strategic Level Working Group. An additional face-to-face meeting for all Tissue Banks and Pathology Tools Workspace participants is being tentatively planned for November.

As with the last edition of the Program Update, in our series of "Meet the Domain Workspaces," the remainder of this issue will be dedicated to the Tissue Banks and Pathology Tools Workspace. The Tissue Banks and Pathology Tools Workspace has had tremendous success in the past few months developing a Project Plan, identifying tools to develop or modify, and defining next steps as they continue to move forward toward their goal of developing an open, interoperable, and modular set of tools to inventory, track, mine, and visualize tissue samples and related information from geographically dispersed repositories.

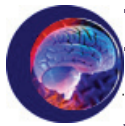
On behalf of NCI, thank you for your enthusiasm and commitment to caBIG.

Sincerely,

Kenneth H. Buetow, Ph.D.
Director

NCI, Center for Bioinformatics

MEET THE DOMAIN WORKSPACES



This Issue— Tissue Banks and Pathology Tools (TBPT) Workspace

The Tissue Banks and Pathology Tools (TBPT) Workspace is one of three Domain Workspaces that was established as part of the caBIG initiative to integrate and/or develop informatics products or solutions that address areas of need for the Cancer Center community. In keeping with the caBIG philosophy of database federation, the goal of the Workspace is to develop an open, interoperable, and modular set of tools to inventory, track, mine, and visualize tissue samples and related information from geographically dispersed repositories. Ultimately this will produce a single point of entrance to federated tissue banks and pathology systems for the research community allowing for a more effective mechanism for researchers to locate and analyze tissue specimens for use in cancer research based on tissue, clinical, and genomic characteristics.

TBPT PARTICIPANTS AND MEETINGS

Cancer Centers with experience in successfully developing tools in this domain are participating in the Workspace 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 have also been invited to participate as early test sites, providing an opportunity to demonstrate real time tool performance. Current participants include institutions with diverse pathology systems:

Developer Institutions:

- Washington University—Siteman
- University of Pittsburgh Medical Center

Adopter Institutions:

- Dartmouth—Norris Cotton
- Thomas Jefferson University—Kimmel
- Northwestern University—Robert H. Lurie
- University of North Carolina—Lineberger
- University of Pennsylvania—Abramson
- Duke University
- Case Western Reserve University – Ireland
- Wake Forest University
- Memorial Sloan Kettering

Working Group Institutions:

- Indiana University
- Jackson Laboratory
- Johns Hopkins—Sidney Kimmel
- University of Alabama at Birmingham
- University of Arizona
- Virginia Commonwealth University—Massey

Participants meet bimonthly via teleconference to discuss current activities and next steps. Artifacts from these meetings can be found on the caBIG website (<http://caBIG.nci.nih.gov>). The first TBPT Workspace face-to-face meeting is tentatively scheduled for November.

To receive information on how to participate in the TBPT Workspace teleconferences, please email the Workspace Coordinator, Greg Eley at eley_greg@bah.com.

CURRENT ACTIVITIES OF THE TBPT WORKSPACE

The TBPT Workspace is currently undertaking three development projects – caTISSUE, caTISSUE Lite and caTIES. As part of these projects, participants are gathering requirements and compiling data about the current systems used by Cancer Centers to inform the development of integrated, grid-enabled tissue banking and pathology informatics tools that meet Cancer Center needs. Where relevant, the group is also studying systems outside of caBIG that may be used as benchmarks for development activities.

Concurrent with these activities, TBPT Workspace participants are working with the Architecture, and Vocabularies and Common Data Elements Workspaces to define data types, adopt data standards and common data elements, and determine a means for communication within the caBIG grid. Ultimately, the goal is to create solutions that adhere to the gold standard for caBIG compatibility. This requires modeling a system to meet both individual Cancer Center needs and cross-cutting caBIG needs to maximize benefits for the cancer research community. Participants are also addressing the need to develop uniform data sharing policies to protect and standardize the sharing of pathology information. These policies will ensure the security and validity of shared information and will facilitate compliance with HIPAA and IRB-related regulatory requirements.

A brief overview of the ongoing development projects in the TBPT Workspace is provided in the following sections.

TOOLS: caTISSUE

A fundamental principle of the caBIG initiative is the integration of information among the cancer research community, leveraging those tools and systems that are already in place, as well as creating new systems to enhance the capabilities and facilitate the compatibility between existing systems. In line with these goals, the TBPT Workspace plans to develop a comprehensive suite of tissue banking tools called caTISSUE. The full suite will provide for the management, query, and retrieval of tissue specimen data and annotations across a federated infrastructure. The cancer research community has a diverse set of requirements ranging from Cancer Centers needing complete end to end systems for the management of tissue and pathology specimens, to those Centers with extensive existing systems interested in providing for interoperability through a modular architecture.

The design and development of caTISSUE will accommodate the diverse range of Cancer Center needs. The most basic form of the caTISSUE suite of tools is caTISSUE Lite, consisting of an easy to use, rapidly deployable system for those Centers who currently do not employ or employ only basic pathology information systems.

caTISSUE Lite provides a functional system that will meet the basic in-house requirements of those users who need to manage specimen databases. It will offer a means for sample identification management, labeling and tracking of specimen samples and specimen aliquots.

TOOLS: caTISSUE

This capability includes mechanisms for ensuring chain of ownership, interfacing with client billing systems, and providing for specimen usage monitoring. caTISSUE Lite will also include the grid services and security services layers which address caBIG cross-cutting requirements. These layers provide for a common messaging system, the usage of standard terminologies and common data elements, employment of information exchange standards, and mechanisms for compliance with HIPAA and related regulatory requirements. The Workspace has set a goal for the first release of caTISSUE Lite within one year.

Concurrent with the development of caTISSUE Lite are a set of additional layers to fill out the caTISSUE suite. These tools will link Centers with existing tissue repositories into a global virtual tissue repository, providing query functionality across the caBIG community. The remaining layers of the caTISSUE suite of tools focus on the integration needs of those Cancer Centers that already employ advanced pathology information systems or that have legacy systems. Existing systems are able to communicate with caTISSUE using Software Development Kits (SDKs) and Advanced Programming Interfaces (APIs). Additionally, caTISSUE will provide modular tools to enhance the processing of tissue and related data, focusing on the ability to associate all manners of annotations to a given specimen. The caTISSUE system will be integrated with tools being developed in other caBIG Workspaces, including integrative cancer research tools and clinical trial management system tools.

TOOLS: caTIES

During initial TBPT Workspace teleconferences, members expressed the need for an automated pipeline to extract data from pathology reports and reformat electronically for use by the cancer research community. Expanding upon the SPIN initiative, Dr. Rebecca Crowley from the University of Pittsburgh Medical Center (UPMC) is leading the Cancer Text Information Extraction System (caTIES) effort. caTIES, an enhancement of SpinTIES, is a general purpose information extraction tool that can be used by caBIG adopters to automate the process of coding, storing, and retrieving data from free-text Pathology Reports. caTIES will leverage standard vocabularies and provide for a grid-enabled tool linking to caTISSUE.

The caTIES system will include:

- A well-designed, front-end for concept-based retrieval from local archives of concept coded pathology reports
- A sufficient data warehouse representation of underlying extracted information to facilitate rapid and accurate query response
- An ongoing contribution to the conceptual operations plan for caBIG based on requirements analysis and initial system development
- A select set of properties, in the NCI Thesaurus, necessary for document classification

The mid-term goal of this project will be to develop a simple interface (to be expanded in later releases) for searching archives of coded pathology reports. At the minimum, the system will allow users to construct queries that permit aggregation of simple demographic

TOOLS: caTIES

information (age, gender), diagnosis, procedure, body-part, TNM Classification, and positive/negative Meta concepts. Initially, the Workspace will create a caBIG compliant application, leveraging existing work to provide a quick deployment for EVS-based concept coding of Pathology Reports. The purpose of this rapid deployment is to quickly empower users with the ability to easily, and automatically concept-code Pathology Reports and make them available as XML documents which can be fed into the architecture adopted by caTISSUE.

To accomplish this task, the Workspace will build on the existing UPMC GATE-based pipeline for concept annotation of Pathology Reports developed previously for the SPIN initiative. The existing system is built on GATE – an open source Java framework for language engineering.

Once the architecture for caBIG has been established, the caTIES query mechanism will be appropriately modified to conform to the accepted architecture. The NCI Thesaurus will be used for assistance in document classification and the CAP-derived CDEs in the caDSR for determining which data to extract. The types of data that caTIES will extract from free-text Pathology Reports include (but are not limited to) the following:

- 1) Body parts present in specimens
 - 2) Surgical procedures performed
 - 3) Diagnoses
 - 4) Findings (present and absent)
 - 5) Values of important attributes (specific to particular Organ/Procedure/ Diagnosis sets) including:
 - TNM classification
 - Tumor grading (Gleason Score, Nottingham Score, etc.)
- Degree of differentiation
 - Presence of absence of lymphatic metastases
 - Presence of absence of marginal involvement
 - Degree of involvement
 - Invasion of adjacent structures
 - Results of immunohistochemical stains and molecular markers reported in free text

The same consideration for caBIG compatibility and compliance with privacy requirements that are guiding the development of caTISSUE are also being observed as part of the caTIES project.

TBPT Workspace – Links to Related Activities in the Cancer Research Community

In the broader cancer research community, parallel and ongoing activities at the NCI and beyond are benefiting the TBPT Workspace and the caBIG community. Many TBPT participants are involved with a variety of ongoing activities outside of caBIG that are providing opportunities for strong integration with existing and evolving tissue and pathology networks. Examples of relevant supporting activities within the cancer research community include:

- Shared Pathology Informatics Network (SPIN): A web-based system capable of searching existing electronic databases to locate human specimens and associated clinical and pathologic data needed for cancer research. The system enables investigators to query databases and receive structured reports with accurate data. The system also establishes data sharing protocols protecting patient privacy and confidentiality.

- National Biospecimen Network (NBN): A national tissue data bank accessible to the cancer research community.
- National Committee for Clinical Laboratory Standards (NCCLS): A global, nonprofit, standards-developing organization that promotes the development and use of voluntary consensus standards and guidelines within the healthcare community.
- Cooperative Prostate Cancer Tissue Resource (CPCTR): A resource that provides researchers with access to primary prostate cancer tissues and their associated clinical and follow-up data. The CPCTR can provide researchers with formalin-fixed, paraffin-embedded primary prostate cancer specimens, and their associated pathologic and clinical information.
- Tissue Array Research Project (TARP): A tool for high throughput screening of multiple tumor tissues using immunohistochemical, in situ, and FISH analyses. The primary objective is to develop and disseminate Multi-Tumor Tissue Microarray slides and related technology to cancer research investigators. TARP is a collaborative effort between The National Cancer Institute and The National Human Genome Research Institute.
- Cooperative Human Tissue Network (CHTN): A network providing biomedical researchers with access to human tissues. The CHTN specializes in the prospective procurement, preservation and distribution of human tissues for research. In addition to normal, benign and malignant tissues, tissues from patients with specific diseases such as ulcerative colitis, a pre-malignant state, are provided.

Next Steps

To continue momentum toward the ultimate vision of providing a rich, comprehensive, interoperable tool set for tissue banks and pathology tools, the Workspace will perform ongoing activities to address all previously identified and any newly identified needs of the community. These needs will be prioritized by Workspace discussions, clarified by white papers, and defined by formal analyses. These activities will ultimately lead to the development of additional tools. Through this continued activity and prioritized development, the Workspace will provide a robust tool set to meet a broad range of Cancer Center needs.

How to get involved with the TBPT Workspace

More specific questions or involvement should be directed to the TBPT Workspace Coordinator, Greg Eley via email at eley_greg@bah.com or by phone at (301) 998-6906.

Track activities on the website at <http://caBIG.nci.nih.gov>.

Watch the Calendar of Events (<http://caBIG.nci.nih.gov/caBIG/calendar>) on the website and What's BIG This Week for upcoming meetings, webcasts, and other events.

caBIG Information and Resources

There are many ways to access additional information and resources on the caBIG initiative and its activities. Following is a list of current resources:

- caBIG Website:

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

- caBIG Workspaces:

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

- caBIG Strategic Level Working Groups:

http://caBIG.nci.nih.gov/caBIG/working_groups

- caBIG Calendar of Events (includes Workspace and Working Group meetings)

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

- Inventory of caBIG infrastructure, applications and datasets:

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

- caBIG Interactive Overview (multi-media presentation on caBIG):

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

- National Cancer Institute Center for Bioinformatics (NCICB) homepage:

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

In addition to these more general resources, a specific caBIG Forum has been established to provide a resource through which all of the caBIG participants can communicate and coordinate with other caBIG members. The forums are readable by anyone, but posting is limited to registered participants.

The caBIG forums can be reached via the caBIG website at <http://caBIG.nci.nih.gov>.

If you do not yet have a login, and you are a registered participant in the caBIG project, you can get one by sending email to Leslie Derr at derrl@mail.nih.gov.

Raising Questions or Concerns

If you are a caBIG participant and you have any questions about caBIG Workspace and Working Group activities, please consider the following avenues to raise your questions:

- The caBIG Forum located on the caBIG website at <http://caBIG.nci.nih.gov>
- Workspace or Working Group Meetings

For specific, directed questions that you feel would be better addressed individually, please feel free to reach out to your respective Workspace/ Working Group Coordinator from the caBIG Project Team, or Mark Adams at adams_mark@bah.com.