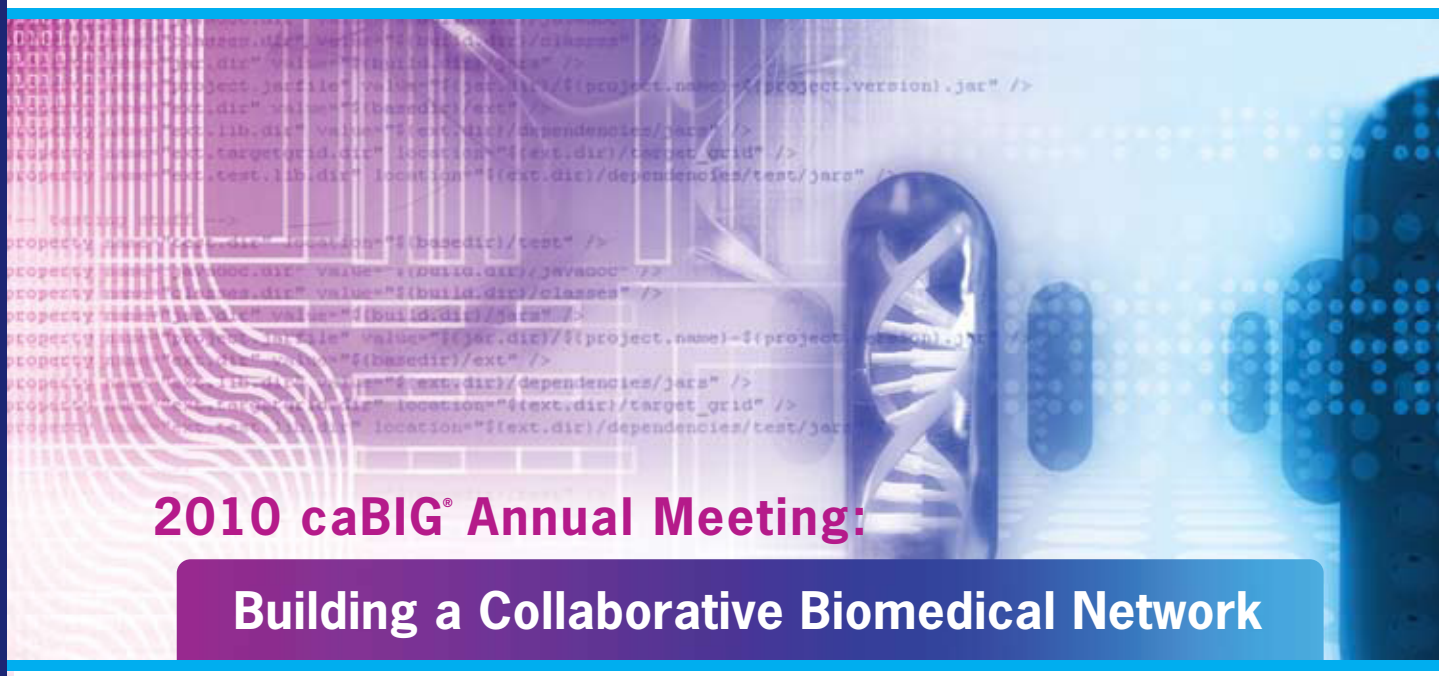
National Cancer Institute



2010 caBIG® Annual Meeting:

Building a Collaborative Biomedical Network

September 13-15, 2010

Marriott Wardman Park Hotel
Washington, D.C.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

Welcome to the 2010 caBIG® Annual Meeting!

This event marks the moment each year when a highly diverse collection of stakeholders – from across scientific disciplines, across sectors, and across geographies – meets to share perspectives about the emerging challenges in biomedicine, and brainstorm together as to how we *as a community* will develop IT-based solutions to accelerate research and improve clinical care.

Looking back for just a moment to the first such conference six years ago, we can say with certainty that the landscape of biomedicine and informatics has changed dramatically. Ideas that were once considered "radical" (i.e., standards-based interoperability, frictionless data liquidity, widespread data-sharing) are now being woven into the fabric of biomedical endeavors. This transformation is gaining momentum exponentially, spurred by multiple trends.

For example, the digitalization of medicine, spurred by the massive investment of the HITECH Act, is now well underway, and clinical care will finally enter the Knowledge Economy. The Health 2.0 "movement," an amalgam of disruptive ideas and disruptive technologies, must be considered in planning infrastructure projects. Consumers and patients, through new online accessible research models, are interacting differently with their healthcare providers and with scientists, presenting themselves as full partners rather than passive recipients. Genetically-specified medicine, heralded since the Mapping of the Human Genome as having huge "potential," is transforming cancer research and becoming part of everyday cancer care.

This trajectory has profound implications for the future of caBIG®. The overarching objective of caBIG® remains before us: to apply information technology to accelerate research and improve clinical care. Hence, caBIG® 2.0 is already evolving, with programs embracing the latest developments in an increasingly mature collection of publicly-available open source infrastructures; expanding our collection of modular and flexible enterprise services; developing new clinical information capabilities; using the emerging capabilities in mobile devices and cloud computing; and above all, making it easier and easier for all stakeholders to exploit caBIG® capabilities.

I am personally looking forward to greeting you at every highlight of this year's meeting: Our formal plenary sessions with illustrious guest speakers, our breakout sessions where we drill deeply into the emerging needs of our many cancer constituencies, and our late-night creativity in designing new capabilities, in what has always been – and I trust will always be – that most exciting intersection where IT innovation meets the best of biology and the cutting edge of clinical care.

With best regards,

Ken Buetow, Ph.D.

Director, Center for Biomedical Informatics and Information Technology
National Cancer Institute

1. caBIG® Program Overview

caBIG® Annual Meeting Overview

- **Are you looking for information about the caBIG® program in general or have meeting-related questions?** Many resources are available at the caBIG® Ask Me Booth. There, you can pick up the caBIG® glossary, which contains basic definitions of commonly used acronyms and concepts, and the annual bibliography, a list of peer-reviewed publications that provides a portrait of ongoing caBIG®-related research. You will also be able to choose from an extensive fact sheet collection that contains not only an overview of the caBIG® program but also detailed information on caBIG tools and grid infrastructure as well as enterprise support services. CBIIT program staff members are at the Ask Me Booth to answer questions as well.

- **Would you like to share experiences and lessons learned with your peers?** Please visit the caBIG® Share Your Story booth at the World's Fair (booth #1). Take advantage of break periods to network and brainstorm with your colleagues and participate in "birds of a feather" evening sessions, which are listed in the agenda.

- **Do you want in-depth information about and hand-on experience with specific caBIG® applications?** Check out the instructor-led interactive sessions and hackathons described in this program book. Also, visit the posters and the tool demos being offered at the World's Fair.

- **Are you seeking comprehensive information about support services available to the caBIG® community?** Refer to the Enterprise Support Network (ESN) section of this program book, and visit the Knowledge Centers and Support Service Providers exhibiting at the World's Fair. Also attend the Vendor Theater adjacent to the World's Fair; the schedule is available in this program book. Fact sheets describing the Enterprise Support Network, the Knowledge Centers, and the Support Service Providers are available at the caBIG® Ask Me Booth.

- **Do you want access to speaker presentations after the Annual Meeting concludes?** All Powerpoint presentations will be made available via the caBIG® Web site (https://cabig.nci.nih.gov/) following the Annual Meeting.

Visit The Share Your Story caBIG® Booth

Tell us about your experience with caBIG®.

caBIG® has always been a collaborative initiative, where experiences and best practices are shared broadly for the benefit of the community. Whether you are just getting started with caBIG®, or have been part of the community for years, we want to hear about your experience.

If you are just getting started, tell us what research challenges you are trying to address. If you are in the process of developing or deploying caBIG® technologies—including software; caGrid; semantics and data-sharing policies—we'd love to hear how this is unfolding in the "real world" for you and your colleagues. And, if you are a long-time veteran, please share the latest ways that caBIG® capabilities are being used at your organization to change the way discovery and translational research are being conducted, clinical care is delivered or "data liquidity" is becoming a reality.

Visit http://cabig.cancer.gov/action/shareyourstory/ to submit information online, or stop by Booth #1 in the World's Fair so we can talk!

Join Us!

Support Service Providers (SSP)
Vendor Theater Schedule
Located in Hall B

Come meet caBIG® Support Service Providers (SSP) at the Annual Meeting! The SSP Vendor Theater at the caBIG® World's Fair provides the opportunity to learn about SSP offerings and meet SSP representatives at a scheduled time outside the formal session schedule. The schedule below list times when SSP organizations will be presenting at the Vendor Theater – stop by to learn about the services and products these support providers bring to the caBIG® community.

SSP Vendor Theater Presentation Schedule

SSP	Presentation Time	SSP's Licensed Support Categories
5AM Solutions	Monday, Sept 13 5:30-6:00 PM	• Deployment Support for caBIG® Software Applications • Adaptation and Enhancement of caBIG® -Compatible Software Applications
LabAnswer	Monday, Sept 13 6:15-6:45 PM	• Deployment Support for caBIG® Software Applications • Help Desk Support • Documentation and Training Materials and Services
HeathCare IT, Inc. (HCIT)	Tuesday, Sept 14 12:30-1:00 PM	• Deployment Support for caBIG® Software Applications
SAIC	Tuesday, Sept 14 1:00-1:30 PM	• Adaptation and Enhancement of caBIG® -Compatible Software Applications
INFOTECHSoft, Inc.	Tuesday, Sept 14 5:30-6:00 PM	• Adaptation and Enhancement of caBIG® -Compatible Software Applications
ESAC, Inc	Tuesday, Sept 14 6:15-6:45 PM	• Adaptation and Enhancement of caBIG® -Compatible Software Applications

2. Agenda

Agenda

caBIG® Annual Meeting: Building a Collaborative Biomedical Network

Monday, September 13, 2010

7:00 am – 5:00 pm	**Registration**
7:00 am – 11:30 am HALL C (EXPO LVL)	**World's Fair Set Up: Posters and Tool Presenters Move-In; Exhibitor Move-In Complete**
8:30 am – 9:00 am THURGOOD (MEZZ LVL)	**Welcome Address** **Presenter:** Kenneth H. Buetow, Ph.D., NCI Associate Director for Bioinformatics and Information Technology; Director, NCI Center for Biomedical Informatics and Information Technology (NCI CBIIT)
9:00 am – 10:00 am THURGOOD (MEZZ LVL)	**Keynote Address: The Vision for Personalized Medicine and the Enabling Role of IT** **Presenter:** Patrick Soon-Shiong, M.D., Abraxis Health
10:00 am – 10:15 am	*Break*
10:15 am – 11:45 am THURGOOD (MEZZ LVL)	**Plenary Session: Enabling Translational Research with caBIG®** Designed for both Newcomers and returning caBIG® community members, this session will provide an overview of how caBIG® is supporting the full life cycle of translational research. It will begin with a short overview of the program and will highlight 3 case studies showing the application of interoperable IT technology in support of translational research. The session will also include a brief overview of the entire meeting with information about other events and onsite resources. Session Chair: Juli Klemm, Ph.D., NCI CBIIT **Presenters:** Ken Buetow, Ph.D., NCI CBIIT, Jinghui Zhang, Ph.D., St. Jude Children's Research Hospital; William Carson, M.D., OSU Comprehensive Cancer Center; Laura Hutchins, M.D., University of Arkansas for Medical Sciences; George Komatsoulis, Ph.D., NCI CBIIT
11:45 pm – 1:00 pm	Lunch (on your own and free time to visit World's Fair)
1:00 pm – 3:00 pm	**CONCURRENT BREAKOUT SESSIONS A - C:**
1:00 pm – 3:00 pm THURGOOD (NORTH/EAST) (MEZZ LVL)	**Session A: Strategic Considerations: Preparing for caBIG® Deployment** This strategic session is designed for those planning or preparing to deploy caBIG® resources and infrastructure. The session will provide an overview of the factors to consider when developing a strategic plan for deploying caBIG® within an organization, and key resources available to those planning these steps. Presentations and discussion will include key factors that help shape deployment goals, different options for connecting with caBIG® and key success factors that have proven to be most important for other organizations in their deployment efforts. Two organizations will present their experiences and perspectives on the process. An overview of training resources and the caBIG® support network will also be presented. Session Chair: George Komatsoulis, Ph.D., NCI CBIIT **Presenters:** Miguel Buddle, M.S., Booz Allen Hamilton; Jennifer Brush, M.S., ScenPro; Jeroen Beliën, Ph.D., VU University Medical Center;

George Komatsoulis, Ph.D., NCI CBIIT; Ken Quinn, R.N., Roswell Park; Jennifer Tucker, Ph.D., OKA

1:00 pm – 3:00 pm
THURGOOD
(WEST/SOUTH)

Session B: Achieving Interoperability with caBIG®: An Introduction to the Services-Aware Interoperability Framework (SAIF) and the Enterprise Conformance and Compliance Framework (ECCF)

For technical audiences that may be new to SAIF, this serves an introduction to SAIF. The session will describe the value proposition for modelers and developers in moving to SAIF, including an overview of the current and planned NCI Enterprise Services Inventory. There will be an overview of the ECCF, which is one of the 4 sub frameworks SAIF is composed of, with an emphasis on transitioning from the current "metals" levels of caBIG® compatibility.

Session Chair: Charlie Mead, M.D., M.Sc., and Caterina Lasome, Ph.D., M.B.A., R.N., CPHIMS, NCI CBIIT
Presenter: Charlie Mead, M.D., M.Sc., NCI CBIIT

1:00 pm – 3:00 pm
WILSON ABC
(MEZZ LVL)

Session C: Poster Discussion Session *(NOTE THIS IS A TIME CHANGE FROM WED. SESSION 20)*

This session will feature eight posters selected for rapid fire presentation, allotting 10 -minutes for each with 5 minutes for presentation and 5 minutes for questions.

Session Chair: Mervi Heiskanen, Ph.D., NCI CBIIT
Presenters: *To be announced.*

3:00 pm – 3:15 pm

Break

3:15 pm – 5:15 pm

CONCURRENT BREAKOUT SESSIONS 1 - 4:

3:15 pm – 5:15 pm
THURGOOD
(NORTH/EAST)
(MEZZ LVL)

Session 1: Supporting Scientific Analysis and Clinical Trials Workflows with caBIG® Tools and Technology

This session will focus on how the tools and resources available from caBIG® are being used to support data integration activities. Capabilities such as federated queries, data warehousing, and use of shared services and hubs enable analysis across multiple data types (including imaging, tissue, pathology, genomic and clinical) and integrated clinical trials workflows using existing applications and databases.

Session Chair: Juli Klemm, Ph.D., NCI CBIIT
Presenters: Eliot Siegel, M.D., Ph.D., University of Maryland School of Medicine; Joseph Jen-Sho Chen, M.D., University of Maryland School of Medicine; John Speakman, NCI CBIIT

3:15 pm – 5:15 pm
WASHINGTON 3 - 4
(EXPO LVL)

Session 2: Deploying caBIG® Technology to Support Biomedical Informatics Needs at Your Organization

This session will continue the discussion presented earlier in Session A and will go a step further to hear from three organizations with diverse workflows and needs that have completed their strategic planning processes. The presenters will discuss their perspective on the drivers and the decisions encountered while planning for caBIG® deployment and implementation.

Session Chair: Miguel Buddle, M.S., Booz Allen Hamilton
Presenters: Jack London, Ph.D., Thomas Jefferson University; David Lloyd Steffen, Ph.D., Baylor College of Medicine; Sorena Nadaf, M.S., M.M.I., University of California at San Francisco

3:15 pm – 5:15 pm
THURGOOD

Session 3: Architecting for Interoperability

Session Chair: Avinash Shanbhag, NCI CBIIT

(WEST/SOUTH)	**Presenters:** *To be announced.*
3:15 pm – 5:15 pm WILSON ABC (MEZZ LVL)	**Session 4: Facilitating Data Sharing through the Data Sharing and Security Framework** This session will focus on real-world application of the DSSF tools to address problems faced by centers deploying caBIG®. It will include a discussion of key data sharing issues and policy questions generally encountered when initiating a caBIG® deployment effort and an overview of the key stakeholders and resources that can be accessed to navigate these issues will occur in the context of these cases. In addition, this session will address legal, regulatory and policy issues faced when enabling data sharing. Session Chair: Wendy Patterson, J.D., NCI TTC **Presenters:** Elaine L. Brock, M.H.S.A., J.D., University of Michigan; Karen J. Maschke, Ph.D., The Hastings Center; Terry Braun, Ph.D., University of Iowa
5:00 pm – 5:15 pm	*Break*
5:15 pm – 7:00 pm HALL C (EXPO LVL)	**WORLD'S FAIR "STAFFED" AND *SSP VENDOR THEATER *(*see Tab 1 for schedule.)*** Exhibits, posters and demonstrations with light refreshments
7:00 pm – 9:00 pm	**ANCILLARY AND BIRDS OF A FEATHER**
7:00 pm – 9:00 pm HALL B (EXPO LVL)	**Birds of a Feather (BOF #1): Global Specimen Identifier (GSID) Service** This session will provide an open forum to discuss the current plans for the Global Specimen identifier, and how it may be used at institutes and at NIH/NCI programs. We welcome community input on how NCI-GSID approach complements your institution's approach and discuss the benefits/challenges in adopting GSID. For more information go to Tissue/Biospecimen KC Forum, see GSID. Session Chair: Ian Fore**,** D.Phil, NCI CBIIT **Presenters:** Tissue Bank/Biospecimen KC and Workspace Representatives
7:00 pm – 9:00 pm WILSON C (MEZZ LVL)	**Imaging Workspace SME Meeting** *Refer to the Imaging Workspace team/lead for more information*.

Tuesday, September 14, 2010

7:00 am –5:00 pm
ATRIUM (EXPO LVL)

Registration

7:00 am – 7:00 pm
HALL C
(EXPO LVL)

World's Fair Exhibition
The exhibit hall will remain open throughout the conference. Dedicated times both Tuesday and Wednesday are set aside when the hall will be staffed by tool and poster presenters and exhibitors.

8:30 am – 9:45 am
THURGOOD
(MEZZ LVL)

Keynote Panel: The Role of Patients and Consumers in New Models of Research
This session will focus on ways that patients and consumers can engage in new ways with biomedical and clinical research through a variety of IT-enabled endeavors. Dr. Love will provide an update on the Army of Women, and discuss how the Health of Women Study serves as a model in cancer that may be applicable to many other disease areas. Mr. Sellers will describe new NCCS programs, highlighting online capabilities that directly address the needs of patients and survivors. Ms. Collyar will provide a perspective on how such transformational activities in research and care are viewed by patients, giving specific examples to enhance understanding and working relationships among all stakeholders in the biomedical ecosystem.

Session Chair: Kenneth Buetow, Ph.D., NCI CBIIT
Presenters: Susan Love, M.D., Susan Love Research Foundation; Thomas P. Sellers, MPA, National Coalition for Cancer Survivorship; Deborah Collyar, PAIR: Patient Advocates In Research and caBIG® Patient Advocate

9:45 am –10:15 am

Break

10:15 am – 12:15 pm

CONCURRENT BREAKOUT SESSIONS 5 – 9:

10:15 am – 12:15 pm
WASHINGTON 1
(EXPO LVL)

Session 5: Integrating Clinical, Imaging and Tissue Annotations with Genomic Data and Medical Images
Hands-on session, seating limited to 50, bring your computer. A limited number of loaner laptops are available in the room.
Using the caIntegrator Translational Research Data Portal this instructor-led session will walk through the process of creating and querying studies on the translational research data portal caIntegrator. Specific example data will come from NCI-sponsored programs, including the NCI Director's Challenge Lung Adenocarcinoma Study and The Cancer Genome Atlas (TCGA) project. Emphasis is on ease of use for integrating data from patient and tissue annotations, gene expression, and medical images.

Session Chair: Mervi Heiskanen, Ph.D., NCI CBIIT
Presenters: Karen A. Ketchum, Ph.D., E-SAC, Inc.

10:15 am – 12:15 pm
WASHINGTON 2
(EXPO LVL)

Session 6: Supporting Translational Research and Care in the Community: Examples from the National Community Cancer Centers Program (NCCCP)
This session will show successful caBIG® implementation at smaller sites and educate caBIG® community about NCCCP program, where collaboration is encouraged. Participants will hear a brief overview of the NCI Community Cancer Center Program (NCCCP) including goals, participants and projects. Speakers will discuss their experiences with implementing caBIG® tools, highlighting the rationales behind decisions to implement caBIG® and the process of implementation, including lessons learned. These sites will discuss a number of unique projects they are undertaking as a result of their NCCCP work.

Session Chair: Brenda Duggan, R.N., NCI CBIIT
Presenters: Brenda Duggan, NCI, CBIIT; Dr. Howard Zaren, Medical Director of the Nancy N. and J.C. Lewis Cancer and Research Pavilion; Beverly Albury, Nancy N. and J.C. Lewis Cancer and Research Pavilion; Joshua Mann, St. Joseph Hospital-Orange

10:15 am – 12:15 pm
HOOVER
(MEZZ LVL)

Session 7: Advancing Population Science and Cancer Control Research Using caBIG® Tools and Infrastructure

This session will provide examples and stories of how caBIG® tools and infrastructure have helped advance population science and cancer control research. Topics covered will include: the Grid-Enabled Measures (GEM) project, a presentation by an end-user on how they are using it and the value it brings to their research; an overview of the infrastructure being used for the PopSciGrid Portal project and how the next version will enable researchers to conduct more interesting analyses; a presentation by Fox Chase Cancer Center in adopting caBIG® to advance population science research.

Session Chair: John Speakman, NCI CBIIT
Presenters: Mike Collins, M.S., Fox Chase Cancer Center; Paul Courtney, Ph.D., SAIC-F; Russ Glasgow, Ph.D., NCI DCCPS; Rick Moser, Ph.D., NCI DCCPS; Eric Ross, Ph.D., Fox Chase Cancer Center

10:15 am – 12:15 pm
THURGOOD
(WEST/SOUTH)
(MEZZ LVL)

Session 8: SAIF and ECCF Implementation in NCI CBIIT

This session is for moderate to highly technical audiences that are already somewhat familiar with the Services Aware Interoperability Framework (SAIF) and the Enterprise Conformance and Compliance Framework (ECCF). It provides a detailed description of the principles of SAIF and ECCF (based on the published NCI CBIIT SAIF Implementation Guide) with emphasis on data types, development philosophy, overview of existing and planned NCI Enterprise services, and mapping current design patterns to SAIF processes.

Session Chairs: Charlie Mead, M.D., M.Sc., and Caterina Lasome, Ph.D., M.B.A., R.N., CPHIMS, NCI CBIIT
Presenters: Christo Andonyadis, D.Sc., NCI CBIIT; Raghu Chintalapati, M.S., M.B.A., Ekagra Software Technologies; Charlie Mead, M.D., M.Sc., NCI CBIIT

10:15 am – 12:15 pm
WASHINGTON 5-6
(EXPO LVL)

Session 9: Securing Data Using caGrid: Use Cases, Policies, and Technologies

This session will provide an overview to developers and legal staff new to caBIG® of the most common security questions and needs that emerge when deploying caBIG® technologies. With specific use cases as the focal point, it will introduce how current and planned caGrid technologies and complementary security policies address these practical needs. Topics may include caGrid installation and configuration behind a firewall; caGrid installation and configuration in a DMZ environment; installing and configuring the Common Security Module; using certificate authority at LOA1; or complying with LOA2 and LOA3 requirements.

Session Chairs: Wendy Patterson, J.D., NCI TTC
Presenters: Braulio Cabral, SAIC-F; Steve Langella, M.S., Inventrio; Marsha Young, J.D., Booz Allen

12:15 pm – 1:30 pm

Lunch (on your own and free time to visit World's Fair and *SSP Vendor Theater)
(*see Tab 1 for schedule.)

1:45 pm – 3:45 pm

CONCURRENT BREAKOUT SESSIONS 10 – 14:

1:45 pm – 3:45 pm
WASHINGTON 1
(EXPO LVL)

Session 10: Finding Data and Analytical Services on caGrid

Hands-on session, seating limited to 50, bring your computer. A limited number of loaner laptops are available in the room.
This instructor-led session will demonstrate the use of the cancer Bench-to-Bedside

(caB2B) application and the caGrid portal to identify services and query data on caGrid. It is aimed at non-technical users and requires no programming skills. Users will come away from the session understanding how to use these tools to find data for their research.

Session Chair: Juli Klemm, Ph.D., NCI CBIIT
Presenters: Mark Adams, Ph.D., Booz Allen Hamilton; Lawrence Brem, M.S., SAIC-F; Jim Humphries, Georgetown University; Konrad Rokicki, M.S., SAIC; Baris E. Suzek, M.S., Georgetown University

1:45 pm – 3:45 pm
WILSON ABC
(MEZZ LVL)

Session 11: Supporting the caBIG® Clinical Research User Community

Session Chair: John Speakman NCI CBIIT
Presenters: Umit Topaloglu, University of Arkansas for Medical Sciences; Robert Annechiarico, Duke Comprehensive Cancer Center
Additional speakers to be announced.

1:45 pm – 3:45 pm
WASHINGTON 3 – 4
(EXPO LVL)

Session 12: Enterprise-wide Data Integration Using caGrid

This session will feature real-world examples of how caGrid can be deployed across an entire organization and connect diverse data.

Session Chair: Avinash Shanbhag, NCI CBIIT
Presenters: Rakesh Nagarajan, M.D., Ph.D. Washington University (WU) School of Medicine; Philip Payne, Ph.D., Ohio State University; John Sandefur, M.B.A., M.S.H.I., University of Alabama at Birmingham Comprehensive Cancer Center

1:45 pm – 3:45 pm
THURGOOD
(WEST/SOUTH)
(MEZZ LVL)

Session 13: caBIG® Semantic Infrastructure v2 Overview

This session is aimed at a moderate to highly technical audience. It will describe the Semantic Infrastructure Version 2 (V2) and its role in enabling computable semantic interoperability in SAIF (Services Aware Interoperability Framework). Participants will receive an overview of key areas expanded upon from V1 to V2, including adoption of a layered and contextual approach to static semantics, and capture of behavioral semantics. Three main functional areas are described: 1. modeling with RIM-based semantics and ISO 21090 datatypes; 2. Form template support; 3. ECCF registry. This is followed by a 25-minute session demonstrating tool prototypes in the above three areas, and a 25-minute overview of new terminology capabilities in LexEVS and CTS2 (Common Terminology Services 2), including value set management and terminology mapping.

Session Chair: Sherri De Coronado, NCI CBIIT and Brian Davis, Ph.D., Booz Allen Hamilton
Presenters: Raghu Chintalapati, M.S., M.B.A., Ekagra Software Technologies; Cecil Lynch, M.D., M.S., University of California, Davis; Charlie Mead, M.D., M.Sc., NCI CBIIT; Craig Stancl, B.S., Mayo Clinic

1:45 pm – 3:45 pm
WASHINGTON 5-6
(EXPO LVL)

Session 14: Regulatory Developments and Practical Barriers to Data Sharing

This session will highlight the range of factors impacting data sharing, with an emphasis on logistical and practical barriers, and the work being done to help facilitate sharing in light of these factors. This session will also explore the particulars of how recent regulatory developments have impacted the data sharing landscape. In particular, this discussion will consider the practical impacts these developments may have on those deploying caBIG® tools, and what can be done at a local level to facilitate the data sharing process.

Session Chair: Wendy Patterson, J.D., NCI TTC
Presenters: Conrade Carl Jaffe, M.D., Boston University; Melissa Markey, J.D., Hall, Render, Killian, Heath & Lyman, PLLC; Rachel Nosowsky, J.D.,; University of California System; Kristen Rosati, J.D., Coppersmith, Schermer & Brockelman, PLC

3:45 pm – 4:00 pm	*Break*
4:00 pm – 4:45 pm THURGOOD (NORTH/EAST)	**AWARDS CEREMONY**
5:00 pm – 7:00 pm HALL C	**WORLD'S FAIR "STAFFED" AND *SSP VENDOR THEATER** ***(*see Tab 1 for schedule.)*** Exhibits, posters and demonstrations with light refreshments
7:00 pm–10:00 pm	Dinner (on your own) and Birds of a Feather and Hack-A-Thon Sessions
7:00 pm – 8:00 pm HALL B	**Birds of a Feather (BOF #2): caRuby: a caBIG® Façade for ETL, Web Services & GUIs** **Presenter:** Fred Loney, Oregon Health & Science University Knight Cancer Institute
8:00 pm–10:00 pm	**CONCURRENT HACK-A-THON SESSIONS 1- 6:** *These sessions are hands-on and computer-based. Please bring your own laptop to participate. Note: *A limited number of loaner laptops are available on a first-come basis, see the caBIG® Booth to reserve.* Seating is limited to 50. For complete session descriptions and requirements see Tab 3c.
8:00 pm–10:00 pm WASHINGTON 1 (EXPO LVL)	**HACK 1: Setting Up a Clinical Trial Using caBIG® Clinical Trials Suite** *Hands-on, bring your computer* Instructor-led workshop on setting up a trial using the caBIG® Clinical Trials Suite and demonstrating data integration and workflow across the applications in the Suite. **Session Leader:** Bill Dyer, Pyramed Research
8:00 pm–10:00 pm WASHINGTON 2 (EXPO LVL)	**HACK 2: Overview of caGrid and Creation of Data Services Using ISO 21090 Data Types** *Hands-on, bring your computer* This session is an overview survey of caGrid capabilities, tools to query the grid, security, and novel applications of caGrid to solve data integration problems. We will also walk users through the creation of a caGrid Data Service for the Patient Outcomes data model. **Session Leaders:** Joe George; The Ohio State University; Justin Permar, Center for IT Innovations in Healthcare; William Stephens, The Ohio State University
8:00 pm–10:00 pm WASHINGTON 3 (EXPO LVL)	**HACK 3: Annotating Microarray Experiments with MAGE-TAB** *Hands-on, bring your computer* This session will cover both avenues for creating MAGE-TAB annotations. Annotare also supports easy incorporation of annotations from relevant biomedical ontologies, a set of standard templates, and a MAGE-TAB validator. Finally, this session will demonstrate how MAGE-TAB files can be imported into and exported from caArray. **Session Leaders:** Rashmi Srinivasa, Ph.D., 5AM Solutions, Inc. and Zhong Li , Ph.D., (MAT KC), Columbia University
8:00 pm–10:00 pm WASHINGTON 4 (EXPO LVL)	**HACK 4: Deploying and Implementing caGrid Services Within Your Organization** *Hands-on, bring your computer* This instructor-led session will draw on the scenarios presented in the morning Grid Security talks to detail how to deploy the caBIG® security applications in one's own organization. **Session Leaders:** Braulio Cabral, Marsha Young; Inventrio: Steve Langella, Shannon Hastings, Scott Oster
8:00 pm–10:00 pm WASHINGTON 5	**HACK 5: Streamlining Data Pipelines in In Silico Research: Creating Basic Research Workflows with Taverna**

(EXPO LVL)

Hands-on, bring your computer Instructor led session showing end-users and bioinformaticians how to create reusable workflows for caBIG® tools using Taverna. Taverna is a suite of tools used to design and execute scientific workflows and aid in silico experimentation.

Session Leaders: Argonne National Lab -The University of Chicago: Ravi Madduri; Dina Sulakhe; Wei Tan, Ph.D.

8:00 pm–10:00 pm
WASHINGTON 6
(EXPO LVL)

HACK 6: caGrid and Cloud Computing: Leveraging Windows Azure for Data and Services

Hands-on, bring your computer In this session, we provide hands-on labs that show the basic capabilities of Windows Azure, which is Microsoft's Cloud platform. We show how to deploy data to Windows Azure, and how to create a caBIG® service in Windows Azure. Developers of all levels and experience are invited, and no previous experience with Windows Azure is necessary (however it is assumed that the attendee is familiar with the basic concepts of cloud computing).

Session Leader: Marty Humphrey, Ph.D., University of Virginia

Wednesday, September 15, 2010

7:00 am – 12 noon	**Registration**
7:00 am – 2:00 pm HALL C	**World's Fair Exhibition**
9:00 am –10:00 am THURGOOD (MEZZ LVL)	**Keynote Address: How Health 2.0 is Changing the Universe and Everything Else** **Matthew Holt, M.A., M.S., Health 2.0, LLC** Session will start with a brief explanation of Health 2.0: What is it? Is it real? And why should anyone care? If he convinces you about any of that, he'll move to the impact of Health 2.0 on patients and providers, data collection and dissemination.
10:00 am – 10:15 am	*Break*
10:15 am – 12:15 pm	**CONCURRENT BREAKOUT SESSIONS 15-19:**
10:15 am – 12:15 pm WASHINGTON 1 (EXPO LVL)	**Session 15: Simplified Consolidation, Updating, and Harmonization of Legacy Biospecimen Data: Using the Bulk Operations Feature of caTissue 1.2** **Hands-on session, seating limited to 50, bring your computer. A limited number of loaner laptops are available in the room.** This session will include hands-on exercises for attendees to follow, from simple basic importing, to more advanced examples using legacy data. Session Chair: Ian Fore, D.Phil., NCI CBIIT **Presenters:** Tissue/Biospecimen Banking and Technology Knowledge Center and the TBPT Workspace
10:15 am – 12:15 pm THURGOOD (WEST/SOUTH) (MEZZ LVL)	**Session 16: In Silico Research Centers of Excellence: Seeing the Future Demise of GBM** This session will cover presentation from three of the ISRCEs and will focus on new understandings of Glioblastoma multiforme based on their work. Session Chair: Mike Berens, Ph.D., Translational Genomics Research Institute (TGen) **Presenters:** Andrea Califano, Ph.D., Columbia University; Daniel Brat, M.D., Ph.D., Emory; Mike Berens, Ph.D., (TGen)
10:15 am – 12:15 pm WILSON ABC	**Session 17: The Role of Standards: Vocabulary and Data Elements to Support Clinical Trials** This session will discuss the critical role of controlled vocabulary/terminology needed to identify, create and register data elements for use in clinical trials. Attendees will have the opportunity to compare and discuss their own CRF variables to the registry of research metadata maintained by the NCI Session Chair: Bill Dyer **Presenters:** Bill Dyer, Pyramed Research; Amy K. Jacobs, M.S.N., R.N.-BC, Lockheed Martin Corporation; Dianne M. Reeves, R.N., M.S.N, NCI CBIIT
10:15 am–12:15 pm WASHINGTON 5 -6	**Session 18: NCI CBIIT Platform/Security/Tooling Roadmap** Session Chair: *To be announced.* **Presenters:** *To be announced.*
10:15 am–12:15 pm WASHINGTON 2	**Session 19: Community Contributed Code: Key Principles and Success Factors**

In this session, members of the caBIG® Community Code Contribution and Integration Workgroup will present the group's work, focusing on key principles and success factors that characterize many open development projects, and which may be effective practices for caBIG® projects as well.

Session Chair: Juli Klemm, Ph.D. NCI CBIIT
Presenters: Mark Adams, Ph.D., Booz Allen Hamilton; Bartley Brown, Ph.D., University of Iowa; Thomas (Tom) Jones, M.D., Tolven, Inc.; Dan Kokotov, M.S., 5AM Solutions; Jim McCusker, Yale School of Medicine

12:15 pm–2:00 pm — Lunch (on your own and free time to visit World's Fair; *Note that World's Fair closes at 2 pm.*)

2:15 pm – 4:15 pm — **CONCURRENT BREAKOUT SESSIONS 20-23:**

This session moved to Mon., Session C

Session 20: Poster Discussion Session (MOVED TO MONDAY 1 PM)

2:15 pm – 4:15 pm
THURGOOD (WEST/SOUTH) (MEZZ LVL)

Session 21: In Silico Research Centers of Excellence: Innovative and Integrative Technologies for Translational Research

Talks from five organizations on the development and use of novel analytical tools and publicly available data to support their research on various forms of cancer. This session will be moderately IT-technical.

Session Chair: Subha Madhavan, Ph.D., Georgetown University
Presenters: Subha Madhavan, Ph.D., Martin Morgan, Ph.D., The Fred Hutchinson Cancer Research Center; Regina Cer, M.S., Uma Mudunuri, The Advanced Biomedical Computing Center, NCI-Frederick; Mariano Alvarez, Ph.D., Columbia University; Jeff Kiefer, Ph.D., The Translational Genomics Research Institute

2:15 pm – 4:15 pm
WILSON ABC

Session 22: Supporting Basic and Clinical Research Across the NCI

Learn about the different research being done in various divisions of the NCI and supported by the caBIG® program from the Cancer Imaging Program, to efforts in the Office of Biorepositories and Biospecimen Research, to Mouse Cancer Models Translating Basic Research to the Clinic.

Session Chair: Ian Fore, D.Phil., NCI CBIIT
Presenters: Sheila Prindiville, M.D., NCI CCR; Cheryl L. Marks, Ph.D. NCI DCB; Laurence P. Clarke, Ph.D., NCI CIP; Carolyn Compton, M.D., Ph.D., NCI OBBR

2:15 pm – 4:15 pm
WASHINGTON 2

Session 23: caBIG® Supporting Integrative Research Beyond Cancer

This session will showcase four projects where caBIG® technology is enabling research beyond cancer from Integrative Informatics Platforms for Clinical and Translational Research, to a NeuroAIDS project, to the Cardiovascular Grid, and also work happening in the Pediatric Heart Network.

Session Chair: Frank White, Ph.D., Feinstein Kean Health Care
Presenters: Philip R.O. Payne, Ph.D., The Ohio State University; Ganesh Shankar, M.S., Indiana University Simon Cancer Center; Raimond L. Winslow, Ph.D., Johns Hopkins University; Anna T. Fernandez, Ph.D., Booz Allen Hamilton

4:15 pm — **MEETING ADJOURNS**

3. Concurrent Sessions

3a. Breakout Sessions

Session and Presentation Descriptions

Monday, September 13, 2010

CONCURRENT BREAKOUT SESSIONS A-C: 1:00 p.m. – 3:00 p.m.

Session A: Strategic Considerations: Preparing for caBIG® Deployment

This strategic session is designed for those planning or preparing to deploy caBIG® resources and infrastructure. The session will provide an overview of the factors to consider when developing a strategic plan for deploying caBIG® within an organization, and key resources available to those planning these steps. First, we will present key factors that help shape deployment goals, including research and data sharing needs, organizational capability and readiness, staffing, and infrastructure. Second, we will outline different options for connecting with caBIG® including deploying caBIG® tools locally, introducing hosted solutions, vendor-provided service options, and internal development efforts. Third, we will highlight key success factors that have proven to be most important for other organizations in their deployment efforts. Next, we will hear from two organizations that have been engaged in deployment projects, to hear their experiences and perspectives on the process. We will close with next steps and resources, including an overview of training resources and the caBIG® support network.

Session Chair: George Komatsoulis, Ph.D., NCI CBIIT
Presenters: Miguel Buddle, M.S., Booz Allen Hamilton; Jennifer Brush, M.S., ScenPro; Jeroen Beliën, M.Sc., Ph.D., VU University Medical Center; George Komatsoulis, Ph.D., NCI CBIIT; Ken Quinn, R.N., Roswell Park Cancer Institute; Jennifer Tucker, Ph.D., OKA

Session B: Achieving Interoperability with caBIG®: An Introduction to the Services-Aware Interoperability Framework (SAIF) and the Enterprise Conformance and Compliance Framework (ECCF)

For technical audiences that may be new to the Services Aware Interoperability Framework (SAIF), this serves as an introduction to SAIF. The session will describe the value proposition for modelers and developers in moving to SAIF, including an overview of the current and planned NCI Enterprise Services Inventory. There will be an overview of the Enterprise Conformance and Compliance Framework (ECCF), which is one of the four sub frameworks SAIF is composed of, with an emphasis on transitioning from the current "metals" levels of caBIG® compatibility.

Session Chairs: Charlie Mead, M.D., M.Sc., NCI CBIIT; Caterina Lasome, Ph.D., M.B.A., CPHIMS, NCI CBIIT
Presenter: Charlie Mead, M.D., M.Sc.

Session C: Poster Discussion Session *(NOTE THIS IS A TIME CHANGE FROM WED. SESSION 20)*

This session will feature eight posters selected for rapid fire presentation, allotting 10-minutes for each with 5 minutes for presentation and 5 minutes for questions.

Session Chair: Mervi Heiskanen, Ph.D., NCI CBIIT
Presenters: *To be announced.*

CONCURRENT BREAKOUT SESSIONS 1 – 4: 3:15 p.m. – 5:15 p.m.

Session 1: Supporting Scientific Analysis and Clinical Trials Workflows with caBIG® Tools and Technology

This session will focus on how the tools and resources available from caBIG® are being used to support data integration activities. Capabilities such as federated queries, data warehousing, and use of shared services and hubs enable analysis across multiple data types (including imaging, tissue, pathology, genomic and clinical) and integrated clinical trials workflows using existing applications and databases.

Session Chair: Juli Klemm, Ph.D., NCI CBIIT

Presenters: Eliot Siegel, M.D., Ph.D., University of Maryland School of Medicine; Joseph Jen-Sho Chen, M.D., University of Maryland School of Medicine; John Speakman, NCI CBIIT

Providing Radiology Observation Data for Genotypic/Phenotypic Analysis in Support of TCGA

This part of the session will focus on how multiple caBIG® technologies are utilized to provide radiology observation data and make it available for cross data analysis.
Presenters: Eliot Siegel, M.D., Ph.D.; Joseph Jen-Sho Chen, M.D.

Clinical Trials Workflow and Data Integration Enabled by the caBIG® Integration Hub Enterprise Service Bus

This part of the session will focus on how the caBIG® Integration Hub supports interoperability by enabling applications – including the caBIG® Clinical Trials Suite – to seamlessly interface with each other and with NCI Enterprise Services.
Presenter: John Speakman

Session 2: Deploying caBIG® Technology to Support Biomedical Informatics Needs at Your Organization

This session will continue the discussion presented earlier in Session A and will go a step further to hear from three organizations with diverse workflows and needs that have completed their strategic planning processes. The presenters will discuss their perspective on the drivers and the decisions encountered while planning for caBIG® deployment and implementation.

Session Chair: Miguel Buddle, M.S., Booz Allen Hamilton
Presenters: David Lloyd Steffen, Ph.D., Baylor College of Medicine; Jack W. London, Ph.D., Thomas Jefferson University; Sorena Nadaf, M.S., M.M.I., University of California at San Francisco

caBIG® at Baylor College of Medicine

The considerations for, planning for and execution of the deployment of three instances of Cancer Biomedical Informatics Grid (caBIG®) tools into production at Baylor College of Medicine will be described; a public instance of caTissue Suite, a private instance of caTissue Suite, and an instance of Central Clinical Participant Registry (C3PR). The three very different motivations for these deployments will be described. The caBIG® Resources that made these deployments possible, including the Data Sharing and Security Framework (DSSF) and the Knowledge Centers, will be referenced. Lessons learned that can inform future decisions about when and how to deploy caBIG® tools will be discussed.
Presenter: David Lloyd Steffen, Ph.D.

Integration of caTissue with Clinical Data in an i2b2 Research Data Mart

At Thomas Jefferson University we have integrated caTissue data with comprehensive de-identified clinical data, including demographics, diagnoses, laboratory values, medications, procedures, and outcomes, from our hospital's clinical data warehouse (CDW) in an i2b2 research data mart (RDM). The informatics for integrating biology and the bedside (i2b2) framework is a widely used approach for integrating clinical and research data. We augmented the i2b2 ontology for the RDM with caTissue concepts. This provides a scientist work flow in which the i2b2 query tool can be used to discover cohorts with the desired phenotype and available specimens, while the caTissue application remains the means of then identifying and ordering the specimens.
Presenter: Jack W. London, Ph.D.

caBIG® and REDCap, CTSA and UCSF Helen Diller Family Comprehensive Cancer Center CRI - Incrementally Advancing Translational Informatics

Presenter: Sorena Nadaf, M.S., M.M.I.

Session 3: Architecting for Interoperability

Session Chair: Avinash Shanbhag, NCI CBIIT
Presenters: To Be Announced

Session 4: Facilitating Data Sharing through the Data Sharing and Security Framework

This session will focus on real-world application of the DSSF tools to address problems faced by centers deploying caBIG®. It will include a discussion of key data sharing issues and policy questions generally encountered when initiating a caBIG® deployment effort, and an overview of the key stakeholders and resources that can be accessed to navigate these issues will occur in the context of these cases. The discussion will also address legal, regulatory and policy issues faced when enabling data sharing.

Session Chair: Wendy Patterson, J.D., NCI CBIIT
Presenters: Elaine L. Brock, M.H.S.A., J.D., University of Michigan; Karen J. Maschke, Ph.D., The Hastings Center; Terry Braun, Ph.D., University of Iowa

BIRDS OF A FEATHER MEETING 7:00 p.m. – 9:00 p.m.

Birds of a Feather (BOF) 1: Global Specimen Identifier (GSID) Service

Applications in clinical and life-science research need to manage information associated with a specimen within an institution as well as across many organizations. For specimen life-cycle tracking, NCI is defining how the Global Specimen Identifier Service (GSID) will be used for various projects, including caHUB and The Cancer Genome Atlas (TCGA) programs. The academic centers, commercial, and hospitals may use global identifiers in their biorepositories to enable tracking and aggregation of clinical and research data. In addition, caTissue, caArray, and other applications will be able to use the GSID to link data from specimens across multiple institutions, through the caGrid. This session will provide an open forum to discuss the current plans for the Global Specimen identifier, and how it may be used at institutes and at NIH/NCI programs. We welcome community input on how NCI-GSID approach complements with your institution's approach and discuss the benefits/challenges in adopting GSID. For up-to-date information on this Birds of a Feather topic, please go to https://cabig-kc.nci.nih.gov/Biospecimen/forums and scroll to the topic title.

Session Chair: Ian Fore, D.Phil, NCI CBIIT
Presenters: Tissue Bank/Biospecimen KC and Workspace Representatives

Tuesday, September 14, 2010

CONCURRENT BREAKOUT SESSIONS 5 – 9: 10:15 a.m. – 12:15 p.m.

Session 5: Integrating Clinical, Imaging and Tissue Annotations with Genomic Data and Medical Images

**This session is computer –based and hands-on with seating limited to 50. We advise arriving early to test equipment. In addition, please be sure to see the software requirements listed in Tab 3b. If you don't have a computer there are a limited number of loaner laptops available in the room. Or, you can participate by listening in.*

Using the caIntegrator Translational Research Data Portal this instructor-led session will walk through the process of creating and querying studies on the translational research data portal caIntegrator. Specific example data will come from NCI-sponsored programs, including the NCI Director's Challenge Lung Adenocarcinoma Study and The Cancer Genome Atlas (TCGA) project. Emphasis is on ease of use for integrating data from patient and tissue annotations, gene expression, and medical images.

Session Chair: Mervi Heiskanen, Ph.D., NCI CBIIT
Presenter: Karen A. Ketchum, Ph.D., E-SAC, Inc.

Session 6: Supporting Translational Research and Care in the Community: Examples from the National Community Cancer Centers Program (NCCCP)

This session will show successful caBIG® implementation at smaller sites and educate caBIG® community about the NCCCP program, where collaboration is encouraged. Participants will hear a brief overview of the NCI Community

Cancer Center Program (NCCCP) including goals, participants and projects. Speakers will discuss their experiences with implementing caBIG® tools, highlighting the rationales behind decisions to implement caBIG® and the process of implementation, including lessons learned. They will also discuss a number of unique projects specific sites are undertaking as a result of their NCCCP work.

Session Chair: Brenda Duggan, R.N., NCI CBIIT
Presenters: Brenda Duggan, R.N., NCI CBIIT; Howard Zaren, M.D., Nancy N. and J.C. Lewis Cancer Center and Research Pavilion; Beverly Albury, Nancy N. and J.C. Lewis Cancer Center and Research Pavilion; Joshua Mann, St. Joseph Hospital-Orange

NCI Community Cancer Center Program (NCCCP): From Pilot to Program
Presenter: Brenda Duggan, R.N

Impact of NCCCP: Broadening Approaches Locally and Nationally
Presenter: Howard Zaren, M.D.

A Unique Approach to Engaging caBIG® Imaging Tools: Challenging the Imaging Community
Presenter: Howard Zaren, M.D.

Mobilizing to Roll-out Multiple caBIG® Tools in Support of a Community Cancer Center: A Template for caBIG® Adoption in the Community
Presenter: Beverly Albury

Novel Deployment Approaches in Leveraging caBIG®
Presenter: Joshua Mann

Session 7: Advancing Population Science and Cancer Control Research Using caBIG® Tools and Infrastructure
This session will provide examples and stories of how caBIG® tools and infrastructure have helped advance population science and cancer control research. It will begin with an overview of the Grid-Enabled Measures (GEM) project, a presentation by an end-user on how they are using it and the value it brings to their research, and description of the next steps, including highlighting the process of taking a measure in GEM to the point of sharing the data on the Grid. A brief overview of the infrastructure being used for the PopSciGrid Portal project will be presented, and how the next version will enable researchers to conduct more interesting analyses. The final presentation in this session will highlight the success of Fox Chase Cancer Center in adopting caBIG® to advance population science research.

Session Chair: John Speakman, NCI CBIIT
Presenters: Michael Collins, M.S., Fox Chase Cancer Center; Paul Courtney, Ph.D., SAIC-F; Russ Glasgow, Ph.D., NCI DCCPS; Rick Moser, Ph.D., NCI DCCPS; Eric Ross, Ph.D., Fox Chase Cancer Center

The Grid-Enabled Measures (GEM) Database: A Science 2.0 Tool to Facilitate the Use of Standardized Measures and Sharing Harmonized Data
The Grid Enabled Measures (GEM) portal provides a forum for a virtual community of researchers to interact with each other using web 2.0 capabilities. The goals of GEM are to promote the use of standardized measures-- tied to theoretically-based constructs --and facilitate the ability to share harmonized data resulting from the use of standardized measures. The second part of the talk will present the efforts of a group to curate measure metadata from selected GEM measures into the caDSR and describe the process used and best practices that were identified. The third part of the talk will include an end-user describing how GEM has supported his research endeavors and describe plans to engage professional organizations in adopting GEM for identifying standardized measures.
Presenters: Michael Collins, M.S.; Russ Glasgow, Ph.D.; Richard P. Moser, Ph.D.

PopSciGrid Portal
Based on a conceptual framework for cyber-enabled collection, harmonization, and analysis of population health data, NCI is developing the PopSciGrid consumer health information portal. The PopSciGrid Portal will demonstrate how tobacco prevalence and policy data can be integrated, visualized, and communicated to help empower

communities and decision-makers. GEM and the PopSciGrid Portal, both of which are integrated with NCI's cancer Biomedical Informatics Grid (caBIG®), focus on supporting greater transparency, scientific collaboration, and community participation in the cancer prevention and control. Grid-enabled applications such as these can target specific public health stakeholders for cancer prevention and control.
Presenter: Paul Courtney, Ph.D.

Development of a Data Integration Platform to Support Population Science and Translational Research

Fox Chase Cancer Center (FCCC) originally developed the Population Research Application Generation Environment (PRESAGE) in 2005 as a toolkit for the efficient deployment of population-based research information systems. In late 2009, FCCC released the second version (2.0) with the goal of providing an integrated platform for translational research. This version provides a portal that integrates data from caBIG® tools with other open source, commercial and internally developed systems. To support efficient, real-time capture of biospecimen data we include a custom interface to the caTissue Suite database. Work continues to capture and manage clinical, analytical and life science data generated by core technology laboratories. In this presentation we will highlight the past, present and future direction of PRESAGE and FCCC's data warehousing efforts for research.
Presenters: Michael Collins, M.S.; Eric Ross, Ph.D.

Session 8: SAIF and ECCF Implementation in NCI CBIIT

This session is for moderate to highly technical audiences that are already somewhat familiar with the Services Aware Interoperability Framework (SAIF) and the Enterprise Conformance and Compliance Framework (ECCF). This session provides a detailed description of the principles of SAIF and ECCF (based on the Published NCI CBIIT SAIF Implementation Guide) with emphasis on data types, development philosophy, overview of existing and planned NCI Enterprise services, and mapping current design patterns to SAIF processes.

Session Chairs: Charlie Mead, M.D., M.Sc., NCI CBIIT; Caterina Lasome, Ph.D., M.B.A., R.N., CPHIMS, NCI CBIIT
Presenters: Christo Andonyadis, D.Sc., NCI CBIIT; Raghu Chintalapati, M.S., M.B.A., Ekagra Software Technologies; Charlie Mead, M.D., M.Sc., NCI CBIIT

Session 9: Securing Data Using caGrid: Use Cases, Policies, and Technologies

This session will open with an overview of the most common security questions and needs that emerge when deploying caBIG® technologies. With these use cases as the focal point, participants will learn how current and planned caGrid technologies and complementary security policies address these practical needs. Topics may include: caGrid installation and configuration behind a firewall; caGrid installation and configuration in a DMZ environment; installing and configuring the Common Security Module (CSM); using certificate authority at LOA1; or complying with LOA2 and LOA3 requirements.

Session Chair: Wendy Patterson, J.D., NCI CBIIT
Presenters: Braulio Cabral, SAIC-F; Steve Langella, M.S., Inventrio; Marsha Young, J.D., Booz Allen Hamilton

CONCURRENT BREAKOUT SESSIONS 10 – 14 1:45 p.m. – 3:45 p.m.

Session 10: Finding Data and Analytical Services on caGrid

**This session is computer –based and hands-on with seating limited to 50. We advise arriving early to test equipment. In addition, please be sure to see the software requirements listed in Tab 3b. If you don't have a computer there are a limited number of loaner laptops available in the room. Or, you can participate by listening in.*
This instructor-led session will demonstrate the use of cancer Bench-2-Bedside (caB2B) and the caGrid portal to identify services and query data on caGrid. It is aimed at non-technical users and requires no programming skills. Users will come away from the session understanding how to use these tools to find data for their research.

Session Chair: Juli Klemm, Ph.D., NCI CBIIT
Presenters: Konrad Rokicki, M.S., SAIC; R. Mark Adams, Ph.D., Booz Allen Hamilton; Lawrence Brem, M.S., SAIC-F; Jim Humphries, Georgetown University; Baris E. Suzek, M.S., Georgetown University

Finding Data and Analytical Services on caGrid: the caGrid iPhone Application

Presenter: Konrad Rokicki, M.S.

Finding Data and Analytical Services on caGrid: the caGrid Portal REST Interface
Presenter: R. Mark Adams, Ph.D.

Finding Data and Analytical Services on caGrid: caGrid Portal
Presenter: Lawrence Brem, M.S.

Finding Data and Analytical Services on caGrid: cancer Bench – to – Bed (caB2B)
Presenters: Jim Humphries; Baris E. Suzek, M.S.

Session 11: Supporting the caBIG® Clinical Research User Community
Session Chair: John Speakman, NCI CBIIT
Presenters: Umit Topaloglu, University of Arkansas for Medical Sciences; Robert Annechiarico, Duke Comprehensive Cancer Center

Session 12: Enterprise-wide Data Integration using caGrid
This session will feature real-world examples of how caGrid can be deployed across an entire organization and connect diverse data types.
Session Chair: Avinash Shanbhag, NCI CBIIT
Presenters: Rakesh Nagarajan, M.D., Ph.D., Washington University; John Sandefur, M.B.A., M.S.H.I., University of Alabama at Birmingham; Philip R.O. Payne, Ph.D., The Ohio State University

Leveraging Service Oriented Architectures and Knowledge Management Methods to Create a Learning Healthcare System
The modern healthcare and life sciences environments have become increasingly data, information, and knowledge intensive. With recent national-scale initiatives intended to advance clinical and translational science, comparative effectiveness research, and personalized healthcare, there is an increased focus on the need to collect, integrate, and reason upon such resources in a timely and effective manner. This presentation will provide an overview of efforts intended to address such needs at The Ohio State University Medical Center, with a particular emphasis on the use of service oriented architectures and comprehensive knowledge management methods. Ultimately, these efforts seek to create an informatics-enabled learning healthcare system, in which every patient encounter is an opportunity to advance the quality and outcomes of biomedical knowledge and practice.
Presenter: Philip R.O. Payne, Ph.D.

Session 13: caBIG® Semantic Infrastructure v2 Overview
For moderate to highly technical audiences, this session describes the Semantic Infrastructure Version 2 (V2) and its role in enabling computable semantic interoperability in SAIF (Services Aware Interoperability Framework). Participants will hear an overview of key areas expanded upon from Version 1(V1) to Version 2 (V2), including adoption of a layered and contextual approach to static semantics, and capture of behavioral semantics. Three main functional areas are described: 1) modeling with RIM-based semantics and ISO 21090 data types; 2) Form template support; 3) ECCF registry. This is followed by a 25-minute session demonstrating tool prototypes in the above three areas, and a 25-minute overview of new terminology capabilities in LexEVS and CTS2 (Common Terminology Services 2), including value set management and terminology mapping.

Session Chairs: Sherri de Coronado, NCI CBIIT; Brian Davis, Ph.D., Booz Allen Hamilton
Presenters: Raghu Chintalapati, M.S., M.B.A., Ekagra Software Technologies; Cecil Lynch, M.D., M.S., University of California, Davis; Charlie Mead, M.D., NCI CBIIT; Craig Stancl, B.S., Mayo Clinic

Session 14: Regulatory Developments and Practical Barriers to Data Sharing
The presenters will highlight the range of factors impacting data sharing, with an emphasis on logistical and practical barriers, and the work being done to help facilitate sharing in light of these factors. This session will also explore the particulars of how recent regulatory developments have impacted the data sharing landscape. In particular, this

discussion will consider the practical impacts these developments may have on those deploying caBIG® tools, and what can be done at a local level to facilitate the data sharing process.

Session Chair: Wendy Patterson, J.D., NCI CBIIT
Presenters: Rachel Nosowsky, J.D., University of California System; Melissa Markey, J.D., Hall, Render, Killian, Heath & Lyman, PLLC; Kristen Rosati, J.D., Coppersmith, Schermer and Brockelman, PLC; Conrade Carl Jaffe, M.D., Boston University

BIRDS OF A FEATHER 7:00 p.m. – 8:00 p.m.

Birds of a Feather (BOF) #2: caRuby: A caBIG® Façade for ETL, Web Services and GUIs

This session organized by Fred Loney will include a brief overview of caRuby and an example of a use case. The presenter will then solicit suggestions and advice from participants. Participants can follow along using their own computer.
Presenter: Fred Loney, Oregon Health & Science University Knight Cancer Institute

HACK-A-THON (HACK) SESSIONS 1 – 6 8:00 p.m. – 10:00 p.m.

**These sessions are computer –based and hands-on sessions, with seating limited to 50. We advise arriving a half hour early to test equipment. In addition, please be sure to see the software requirements listed in Tab 3c. If you don't have a computer there are a limited number of loaner laptops available on a first-come basis; to reserve please inquire at the caBIG® Booth. Or, you may sit in and listen to the session.*

HACK 1: Setting up a Clinical Trial Using caBIG® Clinical Trials Suite

Instructor-led workshop on setting up a trial using the caBIG® Clinical Trials Suite and demonstrating data integration and workflow across the applications in the Suite, *https://wiki.nci.nih.gov/display/Suite*. Applications in the Suite include Patient Registration (C3PR), Study Calendar (PSC), Adverse Event reporting (caAERS), Lab Analysis (Lab Viewer), the caBIG® Integration Hub Enterprise Service Bus for reliable message routing and the Cancer Central Clinical Data Management System (C3D).
Session Leader: Bill Dyer, Pyramed Research

HACK 2: Overview of caGrid and Creation of Data Services using ISO 21090 Data Types

This session is an overview survey of caGrid capabilities, tools to query the grid, security, and novel applications of caGrid to solve data integration problems. The instructors will also walk users through the creation of a caGrid Data Service for the Patient Outcomes data model.
Session Leaders: Joe George, The Ohio State University; Justin Permar, Center for IT Innovations in Healthcare; William Stephens, The Ohio State University

HACK 3: Annotating Microarray Experiments with MAGE-TAB

Annotating experiments is essential to enable unambiguous interpretation and reproducibility of experiments, and also for meaningful search and analysis. MAGE-TAB is an annotation format that allows laboratories to exchange well-annotated biomedical data using a spreadsheet-based paradigm. Several public repositories and tools accept MAGE-TAB annotated data, including caArray, GenePattern, ArrayExpress, Stanford Microarray Database and MeV. MAGE-TAB files can be created from spreadsheet templates, or by using a tool called Annotare that provides intuitive GUI forms to create and modify annotations. This session will cover both avenues for creating MAGE-TAB annotations. Annotare also supports easy incorporation of annotations from relevant biomedical ontologies, a set of standard templates, and a MAGE-TAB validator. Finally, this session will demonstrate how MAGE-TAB files can be imported into and exported from caArray.
Session Leaders: Rashmi Srinivasa, Ph.D., 5AM Solutions, Inc.; Zhong Li, Ph.D., (MAT KC), Columbia University

HACK 4: Deploying and Implementing caGrid Services Within Your Organization

This instructor-led session will draw on the scenarios presented in the morning Grid Security talks to detail how to deploy the caBIG® security applications in one's own organization.

Session Leaders: Braulio Cabral, SAIC-F; Marsha Young, J.D., Booz Allen Hamilton; Steve Langella, Inventrio; Shannon Hastings, Inventrio; Scott Oster, Inventrio

HACK 5: Streamlining Data Pipelines in *In Silico* Research: Creating Basic Research Workflows with Taverna

Instructor-led session showing end-users and bioinformaticians how to create reusable workflows for caBIG® tools using Taverna. Taverna is an open source and domain independent Workflow Management System – a suite of tools used to design and execute scientific workflows and aid *in silico* experimentation. Attendees will implement a plugin for Taverna 2.1 that enables semantic searches for caGrid services described by the caGrid's Index Service and will add them as components to workflows.
Session Leaders: Ravi Madduri, Argonne National Lab -The University of Chicago; Dina Sulakhe, Argonne National Lab -The University of Chicago; Wei Tan, Ph.D., Argonne National Lab -The University of Chicago

HACK 6: caGrid and Cloud Computing: Leveraging Windows Azure for Data and Services

Cloud computing is transforming the IT landscape and will inevitably be important for caBIG®'s future. In this session, we provide hands-on labs that show the basic capabilities of Windows Azure, which is Microsoft's Cloud platform. We show how to deploy data to Windows Azure, and how to create a caBIG® service in Windows Azure. An instructor will be available to answer questions. Developers of all levels and experience are invited, and no previous experience with Windows Azure is necessary (however it is assumed that the attendee is familiar with the basic concepts of cloud computing).
Session Leader: Marty Humphrey, Ph.D., University of Virginia

Wednesday, September 15, 2010

CONCURRENT BREAKOUT SESSIONS 15 – 19: 10:15 a.m. – 12:15 p.m.

Session 15: Simplified Consolidation, Updating, and Harmonization of Legacy Biospecimen Data: Using the Bulk Operations Feature of caTissue 1.2

**This session is computer –based and hands-on with seating limited to 50. We advise arriving early to test equipment. In addition, please be sure to see the software requirements listed in Tab 3b. If you don't have a computer there are a limited number of loaner laptops available in the room. Or, you can participate by listening in.*
Instructor-led session on required steps to use the Bulk Operations Feature available in caTissue 1.2 (release expected in Sept. 2010). This tool can be used to upload legacy or batch data into caTissue Suite. The session will include hands-on exercises for attendees to follow, from simple basic importing to more advanced examples. Example Legacy data will be used in the exercise.

Session Chair: Ian Fore, D.Phil., NCI CBIIT
Presenters: Tissue/Biospecimen Banking and Technology Knowledge Center and the TBPT Workspace

Session 16: In Silico Research Centers of Excellence: Seeing the Future Demise of GBM

This session will cover presentations from three of the ISRCEs and will focus on new understandings of glioblastoma multiforme based on their work.

Session Chair: Mike Berens, Ph.D., Translational Genomics Research Institute
Presenters: Andrea Califano, Ph.D., Columbia University; Daniel Brat, M.D., Ph.D., Emory University; Mike Berens, Ph.D., Translational Genomics Research Institute

Elucidating Mechanisms of Tumorigenesis and Aggressiveness in High-grade Glioma

Presenter: Andrea Califano, Ph.D.

The Emory In Silico Brain Tumor Research Center: Novel Aspects of Glioma Biology Emerge from Applying Advanced Imaging to Molecular Correlates within a caBIG® Environment
Presenter: Daniel Brat, M.D., Ph.D.

Equipping a Flying Leap for Translational Impact Against GBM. Call to Arms.
Presenter: Mike Berens, Ph.D.

Session 17: The Role of Standards: Vocabulary and Data Elements to Support Clinical Trials

This session will discuss the critical role of controlled vocabulary/terminology needed to identify, create and register data elements for use in clinical trials. Use cases, and the global library of standardized CRFs taken from the National Cancer Institute (NCI) Cancer Biomedical Grid (caBIG®) Case Report Form Standardization and Harmonization initiative will be reviewed and analyzed for their ability to collect clinical data that can be both aggregated and analyzed across sites and trials. Attendees will have the opportunity to compare and discuss their own CRF variables to the registry of research metadata maintained by the NCI.

Session Chair: Bill Dyer, Pyramed Research
Presenters: Bill Dyer, Pyramed Research; Amy K. Jacobs, M.S.N., RN-BC, Lockheed Martin Corporation; Dianne M. Reeves, R.N., M.S.N., NCI CBIIT

Session 18: NCI CBIIT Platform/Security/Tooling Roadmap

Session Chair: To Be Announced
Presenters: To Be Announced

Session 19: Community Contributed Code: Key Principles and Success Factors

The caBIG® Community Code Contribution and Integration Workgroup has been working to develop proposals for an enterprise process for evaluating and integrating code from community contributors into the main distribution of existing caBIG® applications. The charter of the group is to identify the range of issues and needs related to code evaluation and integration, and to create a set of proposals for moving forward. In this session, selected group members will present the group's work, focusing on key principles and success factors that characterize many open development projects, and which may be effective practices for caBIG® projects as well. The session will also review the current standard processes for submitting code to the caBIG® program through the Knowledge Centers and the development repository.

Session Chair: Juli Klemm, Ph.D., NCI CBIIT
Presenters: R. Mark Adams, Ph.D., Booz Allen Hamilton; Bartley Brown, Ph.D., University of Iowa; Thomas (Tom) Jones, M.D., Tolven, Inc.; Dan Kokotov, M.S., 5AM Solutions, Inc.; Jim McCusker, M.S., Yale School of Medicine

CONCURRENT BREAKOUT SESSIONS 20 – 23 2:15 p.m. – 4:15 p.m.

Session 20: Poster Presentations—NOTE : MOVED to September 13, 1:00 p.m. (Concurrent Session C)

Session 21: In Silico Research Centers of Excellence: Innovative and Integrative Technologies for Translational Research

Talks from five organizations on the development and use of novel analytical tools and publicly available data to support their research on various forms of cancer. This session will be moderately IT-technical.
Session Chair: Subha Madhavan, Ph.D., Georgetown University
Presenters: Subha Madhavan, Ph.D., Georgetown University; Jeff Kiefer, Ph.D., The Translational Genomics Research Institute; Mariano Alvarez, Ph.D., Columbia University; Martin Morgan, Ph.D., The Fred Hutchinson Cancer

Center; Regina Cer, M.S. The Advanced Biomedical Computing Center, NCI-Frederick; Uma Mudunuri, The Advanced Biomedical Computing Center, NCI-Frederick

Informatics Tools as a Foundation of Rapid Learning Cancer Care and Research Programs
Presenter: Subha Madhavan, Ph.D.

Integrative Workflows for Genomic Analysis and Project Management in a Collaborative Research Environment
Presenter: Jeff Kiefer, Ph.D.

Functional Mining of Human Cancer Interactomes
Presenter: Mariano Alvarez, Ph.D.

Indispensible Technologies for Translational Research: Experiences with iSalmon at the Fred Hutchinson Cancer Research Center
Presenter: Martin Morgan, Ph.D.

bioDBnet: Connecting the Dots in Translational Research. Use Case: Non-B DNA Cancer Research
Presenters: Regina Cer, M.S.; Uma Mudunuri

Session 22: Supporting Basic and Clinical Research Across the NCI
Learn about the different research being done in various divisions of the NCI and supported by the caBIG® program from the Cancer Imaging Program, to efforts in the Office of Biorepositories and Biospecimen Research, to Mouse Cancer Models: Translating Basic Research to the Clinic.

Session Chair: Ian Fore, D.Phil., NCI CBIIT
Presenters: Sheila Prindiville, M.D., NCI CCR; Cheryl L. Marks, Ph.D., NCI DCB; Laurence P. Clark, Ph.D., NCI CIP; Carolyn Compton, M.D., Ph.D. , NCI OBBR

Restructuring the NCI Clinical Trials Enterprise - Clinical Trials Reporting Program (CTRP)
Fulfilling a key recommendation from the NCI Clinical Trials Working Group, all trials conducted at NCI-designated Cancer Centers and all NCI-funded trials conducted elsewhere will be registered in a central database as part of the new Clinical Trials Reporting Program (CTRP). The purpose of the program is to establish a comprehensive database containing regularly updated information on all NCI-funded interventional clinical trials. Grantees will enter specific information about each clinical trial into the database. NCI will use this information to coordinate research efforts to optimize our nation's investment in cancer research. The CTRP central database will leverage the NCI Enterprise Services, deployed on the caBIG® Services-Aware Interoperability Framework (SAIF).
Presenter: Sheila Prindiville, M.D.

Supporting the Mouse Models of Human Cancers Consortium
The mission of the NCI Mouse Models of Human Cancers Consortium is to exploit the ability to alter mouse genes and thereby model cancers (the disease processes) in all organ systems. This program includes the NCI Mouse Repository. The NCI Mouse Repository is an NCI-funded resource for mouse cancer models and associated strains. The repository makes strains available to all members of the scientific community (academic, non-profit, and commercial). NCI Mouse Repository strains are maintained as live colonies or cryoarchived as frozen embryos, depending on demand. Up to three breeder pairs may be ordered from live colonies. Cryoarchived strains are supplied as frozen embryos or recovery of live mice by the NCI Mouse Repository may be requested. This session will describe how caBIG® Imaging, ICR and Tissue technologies have supported the research done within this program over the past several years.
Presenter: Cheryl Marks, Ph.D.

Support Research within the Quantitative Imaging Network
The mission of the Quantitative Imaging network (QIN) is to improve the role of quantitative imaging for clinical decision making in oncology by the development and validation of data acquisition and analysis methods in carefully designed therapy trials, and to create tools to improve the measurement of response to drug or radiation therapy in

order to predict outcome better and thus tailor treatment to individual patients. Investigators are expected to have the ability to share software and related statistical tools that are designed to validate clinical decision tools for the range of modalities. The caBIG® Imaging technologies can enable that sharing of software and imaging data. The caBIG® ICR technologies can enable the integration of the imaging data with other data types for correlative analysis.
Presenter: Laurence P. Clarke, Ph.D.

The Cancer Human Biobank: A National Repository for Human Biospecimens and Data Associated with Them to Support Cancer Research

The need for high-quality human biospecimens exists throughout the cancer research and product development communities. In response to this need, the Office of Biorepositories and Biospecimen Research (OBBR) is launching the cancer Human Biobank (caHUB), the first national biobank of standardized human benchmark specimens for research. No comparable resource exists at this time. Through a caBIG®-compliant public network, caHUB would make specimens and data available to a broad research community, including academia, government, private foundations, and biotechnology and pharmaceutical industries, and enable new collaborations among researchers in all areas of investigation, thereby accelerating the pace of discovery and innovation.
Presenter: Carolyn Compton, M.D., Ph.D.

Session 23: caBIG® Supporting Integrative Research Beyond Cancer

This session will showcase four projects where caBIG® technology is enabling research beyond cancer from Integrative Informatics Platforms for Clinical and Translational Research, to a Neur-AIDS project, to the Cardiovascular Grid, and also work happening in the Pediatric Heart Network.

Session Chair: Frank White, Ph.D., Feinstein Kean Health Care
Presenters: Philip R.O. Payne, Ph.D., The Ohio State University; Ganesh Shankar, M.S., Indiana University Simon Cancer Center; Raimond L. Winslow, Ph.D., The Johns Hopkins University; Anna T. Fernandez, Ph.D., Booz Allen Hamilton

Integrative Informatics Platforms for Clinical and Translational Research: caBIG® Beyond Cancer

The biomedical informatics program of OSU's CTSA-funded Center for Clinical and Translational Science (CCTS) regularly engages in projects that seek to adopt and adapt caBIG® technologies for use in domains beyond cancer. As an example of such efforts, the OSU CCTS is actively implementing an integrative informatics platform, which leverages both caGrid and caGrid Portal, and is intended to link phenotypic and bio-molecular data that can characterize maternal-fetal dyads in order to satisfy a variety of clinical and outcomes research related use cases. This project requires the federation of data repositories at both OSU Medical Center and Nationwide Children's Hospital, as well as the provision of semantic reasoning and end-user presentation models that enable the design and execution of distributed data queries. This presentation will summarize the technical, socio-cultural and workflow implications of such efforts, with a particular emphasis on the design patterns necessary to enable the adaptation of caBIG® technologies for diverse application domains.
Presenter: Philip R.O. Payne, Ph.D.

Using caBIG® Tools in NeuroAIDS Research

The applications produced by the Cancer Biomedical Informatics Grid initiative can be used in areas of research beyond cancer. We have deployed a number of these applications for NeuroAIDS research in the National NeuroAIDS Tissue Consortium (NNTC). The Consortium is using caArray to provide a mechanism for collaboration and data sharing for microarray data, NBIA as a repository of Magnetic Resonance Images and caIntegrator2 as a study portal for integrated microarray, clinical data, and images. The Consortium is also utilizing caTissue for storing tissue specimen information. These specimen data are integrated with their associated tissue microarray images using a non-caBIG® application called TMAj. Leveraging caBIG® tools has dramatically increased the informatics capabilities of the NNTC, within a very short time, while incorporating data standards.
Presenter: Ganesh Shankar, M.S.

Update on the Cardiovascular Research Grid Project

Translational cardiovascular research now involves the collection and analysis of multiple data sets (SNP, mRNA expression, protein expression, imaging, ECG, clinical, etc.) from large cohorts. The CardioVascular Research Grid (CVRG) Project is creating a suite of software tools and applications that enable sharing and analysis of cardiovascular data as part of multi-institutional research projects. To do this, the CVRG team has defined a set of Driving Biomedical Projects that help identify the informatics needs of the cardiovascular research community. The CVRG team then develops software to meet these needs. The project is supported by the National Heart Lung &

Blood Institute. It is an inter-disciplinary, collaborative effort bringing together biomedical and software systems researchers. The project is based at the Institute for Computational Medicine at The Johns Hopkins University, in collaboration with the Center for Comprehensive Informatics at Emory University, the Image Lab at Wake Forest University and the Computation Institute at The University of Chicago. The software architecture of the CVRG builds on and extends software systems developed by the participants of the project at these institutions as well as those developed in the cancer Biomedical Informatics Grid (caBIG®) and Biomedical Informatics Research Network (BIRN) projects.
Presenter: Raimond L. Winslow, Ph.D.

caBIG® and Open-Source Technology to Address Image-Based Clinical Study Management in the NHLBI Pediatric Heart Network

The National Heart, Lung, and Blood Institute's Pediatric Heart Network (PHN) started in 2001 as a collaboration of medical centers to conduct nationwide research studies in children with heart disease. We describe our efforts with PHN to identify their requirements for a web-portal based on open-source tools. This portal may be used for their image-based, multi-center clinical trials to track the status of image submissions from clinical sites to the data coordinating center and to core labs, and provide access to analysis results. An identified need is to provide a way to track a center's submitted images through quality control steps and data processing. Identification and assessment of caBIG® and other technologies was conducted, and resulted in a prototype web portal. Supported with funds from NHLBI-NIH Grant No. U01HL068270.
Presenter: Anna T. Fernandez, Ph.D.

3b. Hands-On Sessions

Hands-on Sessions

There will be three hands-on sessions – two on Tuesday and one on Wednesday. Seating is limited to 50 participants. Please see the enclosed pages for more information, system/computer requirements, level of difficulty, session objectives and more.

If you have any questions please inquire at the caBIG® Ask Me Booth. We STRONGLY suggest arriving 15 -20 minutes early to test computer equipment and wireless Internet.

If you do not have a computer we have a very small number of loaner laptops for participants to share/borrow. Please inquire at the caBIG® Booth. If you are without a computer you are welcome to sit in and listen.

Location: *Please see the agenda for the corresponding session and room.*

Hands-on Sessions

Tuesday, September 14, 10:15 am
Session 5: Integrating Clinical, Imaging and Tissue Annotations with Genomic Data and Medical Images

Using the caIntegrator Translational Research Data Portal this instructor-led session will walk through the process of creating and querying studies on the translational research data portal caIntegrator. Specific example data will come from NCI-sponsored programs, including the NCI Director's Challenge Lung Adenocarcinoma Study and The Cancer Genome Atlas (TCGA) project. Emphasis is on ease of use for integrating data from patient and tissue annotations, gene expression, and medical images.

Session Chair: Mervi Heiskanen, Ph.D., NCI CBIIT
Presenters: Karen A. Ketchum, Ph.D., E-SAC, Inc.

Computer Requirements:
1. IE 7 and above or FireFox 3 and above should suffice

General Information (For your reference only, no action is needed.)
1. Product Description: https://cabig.nci.nih.gov/tools/caIntegrator2
2. Documentation:https://wiki.nci.nih.gov/display/caIntegrator2/caIntegrator2+Wiki

Tuesday, September 14, 1:45 pm
Session 10: Finding Data and Analytical Services on caGrid

This instructor-led session featuring four presentations will demonstrate the use of cancer Bench-2-Bedside (caB2B) and the caGrid portal to identify services and query data on caGrid. It is aimed at non-technical users and requires no programming skills. Users will come away from the session understanding how to use these tools to find data for their research.

Session Chair: Juli Klemm, Ph.D., NCI CBIIT
Presenters: Konrad Rokicki, M.S., SAIC; R. Mark Adams, Ph.D., Booz Allen Hamilton; Lawrence Brem, M.S., SAIC-F; Jim Humphries, Georgetown University; Baris E. Suzek, M.S., Georgetown University

(1) Finding Data and Analytical Services on caGrd: the caGrid iPhone Application
Presenter: Konrad Rokicki, M.S.

(2) Finding Data and Analytical Services on caGrid: the caGrid Portal REST Interface
Presenter: R. Mark Adams, Ph.D.

(3) Finding Data and Analytical Services on caGrid: caGrid Portal
Presenter: Lawrence Brem, M.S.

(4) Finding Data and Analytical Services on caGrid: cancer Bench – to – Bed (caB2B)
Presenters: Jim Humphries; Baris E. Suzek, M.S.

Computer Requirements:
1. IE 7 and above or FireFox 3 and above should suffice

Wednesday, September 15, 10:15 am
Session 15: Simplified Consolidation, Updating, and Harmonization of Legacy Biospecimen Data: Using the Bulk Operations Feature of caTissue 1.2

Instructor-led session on required steps to use the Bulk Operations Feature available in caTissue 1.2 (release expected in Sept. 2010). This tool can be used to upload legacy or batch data into caTissue Suite. The session will include hands-on exercises for attendees to follow, from simple basic importing to more advanced examples. Example Legacy data will be used in the exercise.

Session Chair: Ian Fore, D.Phil., NCI CBIIT
Presenters: Tissue/Biospecimen Banking and Technology Knowledge Center and the TBPT Workspace

Computer Requirements:
1. IE 7 and above or FireFox 3 and above should suffice

3c. “Hack-a-thon”

Hack-a-thon Sessions

A tradition of the caBIG® Annual Meeting, the Hack-a-thon sessions are another opportunity for hands-on and computer-based learning.

We are offering six sessions which will take place on Tuesday evening from 8:00 pm to 10:00 pm. The maximum seating for each is 50. Please see the enclosed pages for more information, system/computer requirements, session level of difficulty, and learning objectives.

If you have any questions please inquire at the caBIG® Ask Me Booth. We STRONGLY suggest arriving 30 minutes in advance to test computer equipment and wireless Internet.

We have a very small number of loaner laptops for participants to share/borrow, for more information see the caBIG® Booth. If you are without a computer you are welcome to sit in and listen.

Location: Washington Rooms 1 thru 6

See the agenda for the corresponding session and room.

Hack-a-thon Sessions

Hack-a-thon (1) – Setting up a Clinical Trial Using caBIG® Clinical Trials Suite

Instructor-led workshop on setting up a trial using the caBIG® Clinical Trials Suite and demonstrating data integration and workflow across the applications in the Suite, https://wiki.nci.nih.gov/display/Suite. Applications in the Suite include Patient Registration (C3PR), Study Calendar (PSC), Adverse Event reporting (caAERS), Lab Analysis (Lab Viewer), the caBIG® Integration Hub Enterprise Service Bus for reliable message routing and the Cancer Central Clinical Data Management System (C3D).

Presenter: Bill Dyer

1. Computer Requirements:
None provided.

Hack-a-thon (2) – Overview of caGrid and Creation of Data Services Using ISO 21090 Data Types

This session is an overview survey of caGrid capabilities, tools to query the grid, security, and novel applications of caGrid to solve data integration problems. The instructors will also walk users through the creation of a caGrid Data Service for the Patient Outcomes data model.

Presenters: Joe George, Justin Permar, William Stephens

1. Computer Requirements:
   a. IE 7 and above or FireFox 3 and above should suffice
   b. At least 5 GB of free memory
   c. JAVA 6 http://java.com/en/download/inc/windows_upgrade_ie.jsp

General Information (For reference purposes, no action is needed)
   a. caGrid Data Services http://cagrid.org/display/dataservices/Home
   b. ISO 21090 https://wiki.nci.nih.gov/display/ISO21090/NCI+CBIIT+ISO+21090+Support+in+caCORE+SDK%2C+caAdapter%2C+and+caGrid

Hack-a-thon (3) – Annotating Microarray Experiments with MAGE-TAB

Annotating experiments is essential to enable unambiguous interpretation and reproducibility of experiments, and also for meaningful search and analysis. MAGE-TAB is an annotation format that allows laboratories to exchange well-annotated biomedical data using a spreadsheet-based paradigm. Several public repositories and tools accept MAGE-TAB annotated data, including caArray, GenePattern, ArrayExpress, Stanford Microarray Database and MeV. MAGE-TAB files can be created from spreadsheet templates, or by using a tool called Annotare that provides intuitive GUI forms to create and modify annotations. This session will cover both avenues for creating MAGE-TAB annotations. Annotare also supports easy incorporation of annotations from relevant

biomedical ontologies, a set of standard templates, and a MAGE-TAB validator. Finally, this session will demonstrate how MAGE-TAB files can be imported into and exported from caArray.

Presenters: Rashmi Srinivasa, Zhong Li

1. Computer Requirements
    a. MAC OS X (10.5 or greater) OR Windows (XP, Vista, 7)
    b. At least 190M disk space for Mac OR 70M for Windows
    c. JAVA http://java.com/en/download/inc/windows_upgrade_ie.jsp
2. Install on Windows
    a. Uninstall any older version of Annotare via the Control Panel and then delete the Annotare folder under "Program Files".
    b. Download the Installer bundled with/without the JRE (Java Runtime Environment):
    c. http://annotare.googlecode.com/files/Annotare-v1.1.1-jre.exe (Recommended)
    d. http://annotare.googlecode.com/files/Annotare-v1.1.1.exe (You must install JRE 1.5 or 1.6 yourself)
    e. Double-click on the exe and follow the steps.
    f. Go to Programs -> Annotare to start the application.
3. Install on MAC OS X
    a. Uninstall any older version of Annotare by removing it from /Applications.
    b. Use the Software Update feature (available from the Apple menu) to get the latest version of Java (if you don't already have 1.5 or higher)
    c. Download the installer: http://annotare.googlecode.com/files/Annotare-v1.1.1.pkg.zip
    d. Double-click to unzip, then double-click on the .pkg to install.
    e. Click on /Applications/Annotare to start the application.

Hack-a-thon (4) – Deploying and Implementing caGrid Security Services Within Your Organization

This instructor-led session will draw on the scenarios presented in the morning Grid Security talks to detail how to deploy the caBIG® security applications in one's own organization.

Presenters: Braulio Cabral, Marsha Young, Stephen Langella, Shannon Hastings, Scott Oster

1. Computer Requirements:
    a. IE 7 and Above or FireFox 3 and above should suffice
2. Pre-Material:
    a. This session will mostly be a demonstration and information session, however, everything thing we go over will be available with complete "do it yourself" steps on the cagrid.org site at: http://www.cagrid.org/display/cagrid13tutorials/Deploying+a+Secure+caGrid+Service
    b. Folks participating can follow along by downloading the required tooling (we are still putting together) from that site, prior to attending the tutorial. Likewise folks can download everything after the hackathon and complete everything using the step by step instructions that will be provided on that page.

Hack-a-thon (5) – Streamlining Data Pipelines in In Silico Research: Creating Basic Research Workflows with Taverna

Instructor-led session showing end-users and bioinformaticians how to create reusable workflows for caBIG® tools using Taverna. Taverna is an open source and domain independent Workflow Management System – a suite of tools used to design and execute scientific workflows and aid *in silico* experimentation. Attendees will implement a plugin for Taverna 2.1 that enables semantic searches for caGrid services described by the caGrid's Index Service and will add them as components to workflows.

Presenters: Wei Tan, Dina Sulakhe, Ravi Madduri

1. Computer Requirements:
    a. IE 7 and Above or FireFox 3 and above should suffice
    b. JAVA http://java.com/en/download/inc/windows_upgrade_ie.jsp
    c. Please see pages "Taverna" that appear after Hackathon6.

Hack-a-thon (6) – caGrid and Cloud Computing: Leveraging Windows Azure for Data and Services

Cloud computing is transforming the IT landscape and will inevitably be important for caBIG®'s future. In this session, we provide hands-on labs that show the basic capabilities of Windows Azure, which is Microsoft's Cloud platform. We show how to deploy data to Windows Azure, and how to create a caBIG® service in Windows Azure. An instructor will be available to answer questions. Developers of all levels and experience are invited, and no previous experience with Windows Azure is necessary (however it is assumed that the attendee is familiar with the basic concepts of cloud computing).

Presenter: Marty Humphrey

1. Computer Requirements:
    a. IE 7 and above or FireFox 3 and above should suffice
2. Pre-Material:
    a. http://www.cs.virginia.edu/~humphrey/caBIG/ contains the three labs from last year's hack-a-thon. Interested attendees of this year's Hackathonshould consider looking at this web page to both [a] go though one or more of these labs themselves if they wish (we will have different labs this year) and [b] see the style of the labs (which will be similar this year)

Please READ and follow steps to participate in Streamlining Data Pipelines In Silico Research: Creating Basic Research Workflows with Taverna (Hackathon 5)

Installation Guide to caGrid Workflow Suite and Taverna 2.1.2
Streamlining Data Pipelines in In Silico Research: Creating Basic Research Workflows with Taverna

Overview of caGrid Workflow Suite and Taverna 2.1.2
The Taverna Workbench allows users to construct complex workflows consisting of multiple types of components. Each type of component is called an "Activity". Once combined in a workflow, these components, which may be located on different machines, are orchestrated by Taverna which then gathers the results and displays them in the workbench interface.
The 2.1.2 release of Taverna supports many types of activities, including: apiconsumer activity, beanshell activity, biomart activity, biomody activity, java activity, soaplab activity, and wsdl activity, among others.

The caGrid Workflow team has built caGrid Workflow Suite as a set of Taverna plug-ins that facilitates the discovery, invocation, and execution of caGrid services through the Taverna Workbench. caGrid Workflow Suite contains these plug-ins:

- cagrid-Activity: This plug-in is used to query caGrid services from the caGrid Index Service, invoke the service, manage security configuration, etc This allows you to easily find the available caGrid services and select and leverage them as needed for your workflow.
- caGrid-Service: This workflow-execution plug-in passes a workflow definition file, including the appropriate inputs, to a generic caGrid service, where the workflow is then executed. This is useful in instances where the workflow is long-running and relies on constant access to the Grid to execute properly.
- caGrid-Transfer-Activity: leverages the caGrid transfer utility to move files between services and clients using HTTP protocol, without embedding them in SOAP messages.
- CQL-Builder-Activity: assists users to build a CQL clause through GUI.
- CDS-Activity: assists users to delegate credentials to a 3rd party.
- Taverna 2.1.2 also supports Grid services built using WSRF specifications.
- Getting Started with Taverna 2.1.2
- The first step to using Taverna is to install Taverna 2.1.2 (also referred to as the Taverna Workbench).

Contact for questions:
caGrid Workflow Team:
Ravi Madduri: madduri@mcs.anl.gov
Wei Tan: wtan@mcs.anl.gov
Dinanath Sulakhe: sulakhe@mcs.anl.gov

Before installing the program, you should review the system requirements for installation. That information is available at the following URL: http://www.mygrid.org.uk/tools/taverna/taverna-requirements/.

- For the most part, Taverna 2.1.2 works with Java 1.5 or higher.
  The Taverna 2.1.2 Workbench program is available for download from the following URL: https://launchpad.net/taverna/+milestone/2.1.2. The installation instructions can be found at http://www.taverna.org.uk/download/workbench/.

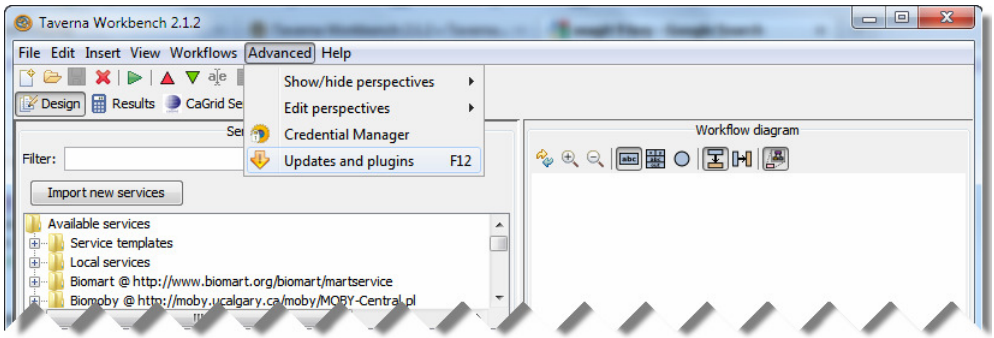
- If you are using Linux you must install the graphviz package so that Taverna can find the "dot" executable. The graphviz package can be found at the following URL: http://www.research.att.com/sw/tools/graphviz/download.html.

If this is the first time you are using Taverna 2.1.2, be advised that it may take some time to initiate the program, since it will need to fetch some components from web.
If you have used previous versions of Taverna (Taverna 1.x), you may want to review the Quick Guide for Taverna 1 Users. This guide provides a basic feature comparison between Taverna 1.x and 2.x and identifies those features that are not provided in 2.x. The quick guide also lists the restrictions for upgrading your existing 1.x workflows to 2.x workflows. The quick guide is located at the following URL: http://www.mygrid.org.uk/tools/taverna/taverna-2-0/taverna-2-0-documentation/taverna-2-0-help/quick-guide/.
You can find a more complete version of the Taverna 2.x user documentation at the following URL: http://www.taverna.org.uk/documentation/taverna-2-x/.

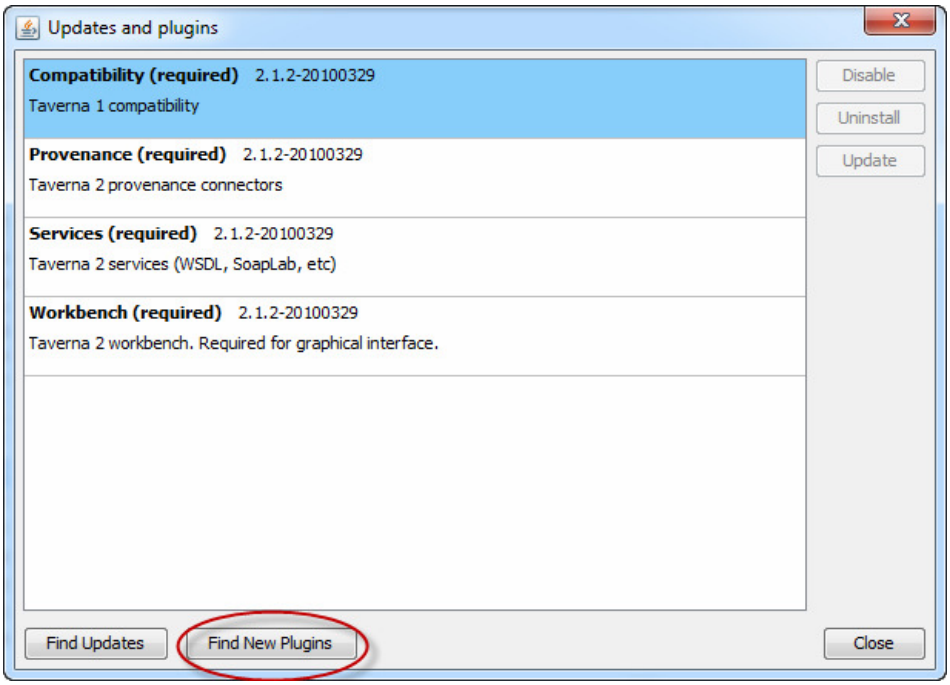
Download the caGrid Workflow Suite

caGrid has wrapped all caGrid related plug-ins into one called "caGrid Workflow Suite". In order to use this caGrid plug-in, you must download them from the Taverna interface.
To download the caGrid Workflow Suite:

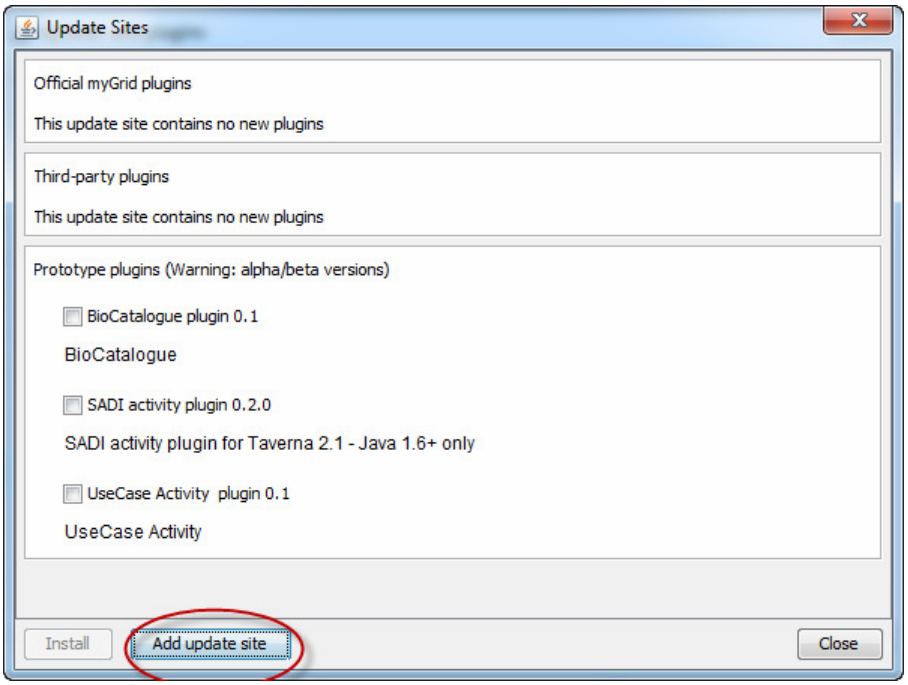
1. Start up the Taverna 2.1.2 Workbench.
2. From the **Advanced** option on the main menu, select **Updates and plug-ins**.



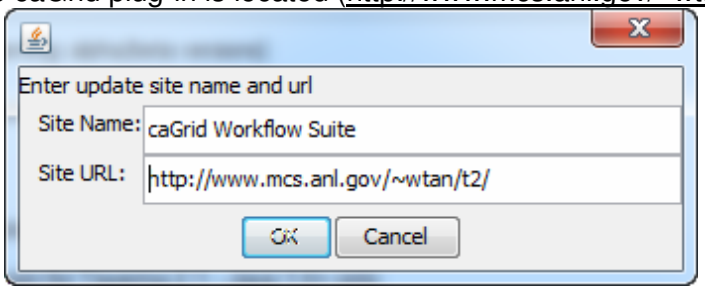
3. In the Plug-in Manager dialog box that appears, click **Find New Plug-ins** button located at the bottom of the dialog box.



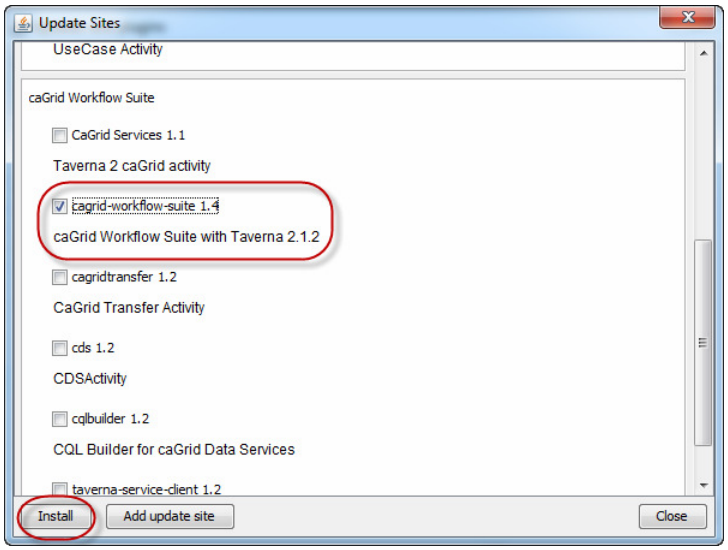
4. In the **Plug-in Sites** dialog box that appears, click **Add Plug-in Site** located at the bottom of the dialog box.



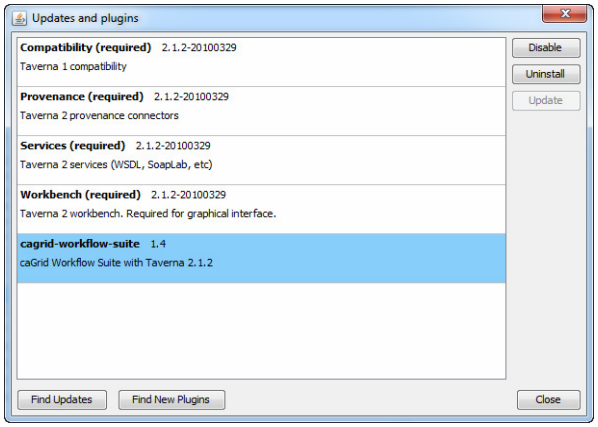
5. A dialog box appears that allows you to identify the site from which you want to download the plug-in(s). Enter the name (e.g. caGrid Workflow Suite) and the URL of the site where the caGrid plug-in is located (http://www.mcs.anl.gov/~wtan/t2/) and click **OK**.



6. When the Plug-in Sites dialog box reappears, it should show the Taverna-caGrid plug-in found at this site. Make sure the checkbox for the plug-in is checked and click **Install**.

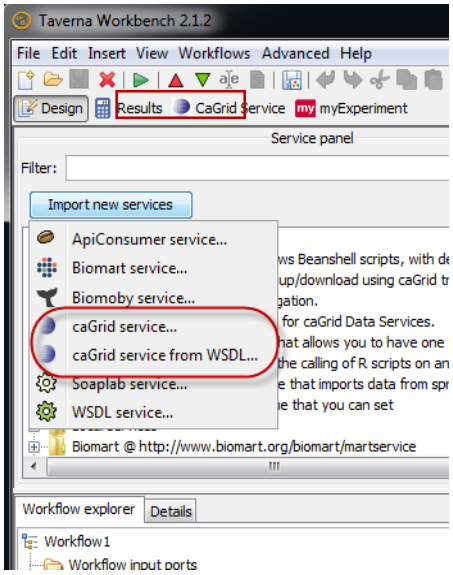


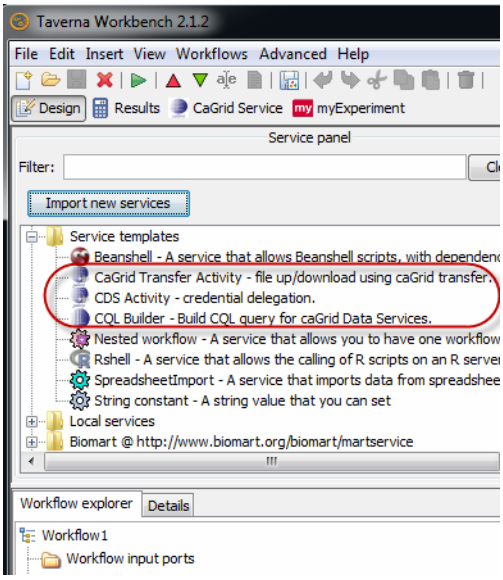
NOTE: This step may take some time depending on the network speed, while the plug-in is downloaded and installed.



7. After the plug-in installation is complete, restart Taverna Workbench to allow the plug-in installation to take effect.

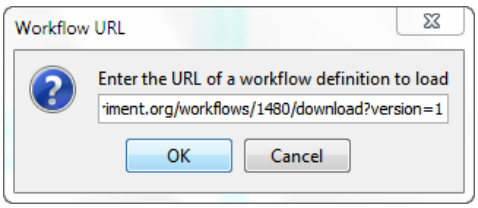
   When the Taverna Workbench reappears, you can check the interface to verify the installation of the caGrid Workflow Suite went correctly. The caGrid workflow suite currently contains five plug-ins, i.e., cagrid-activity (**caGrid service…** and **caGrid service from WSDL…**) for service discovery, invocation and security enforcement; cql-builder (**CQL Builder**) for visualized construction of CQL clause to query data services; cagrid-transfer-activity (**CaGrid Transfer Activity**) for file transfers between clients and services; cds-activity (**CDS Activity**) for credential delegation; caGrid-service (CaGrid Service) for workflow execution in a caGrid service.



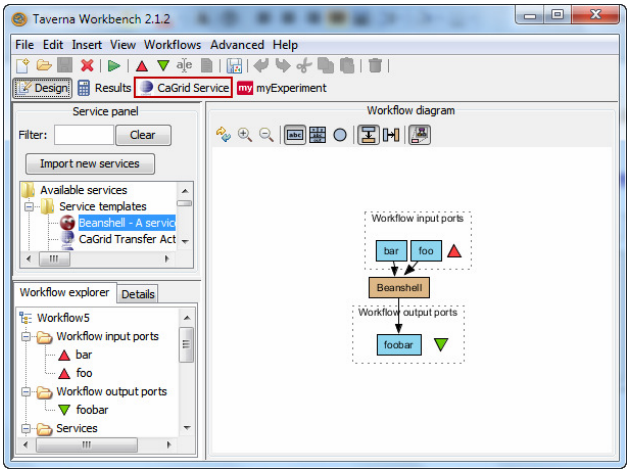


Validation of the installation

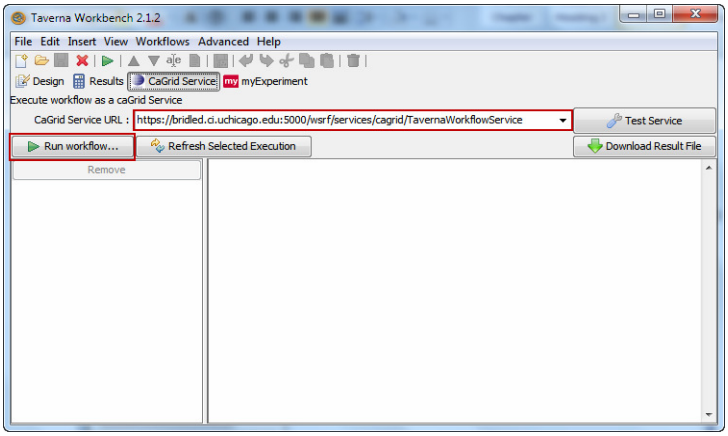
This step validates the connection with caGrid Workflow Service by submitting a helloworld workflow to it.
From the workbench, select **File > Open workflow location**, in the popped up dialog, give the URL below:
http://www.myexperiment.org/workflows/1480/download?version=1



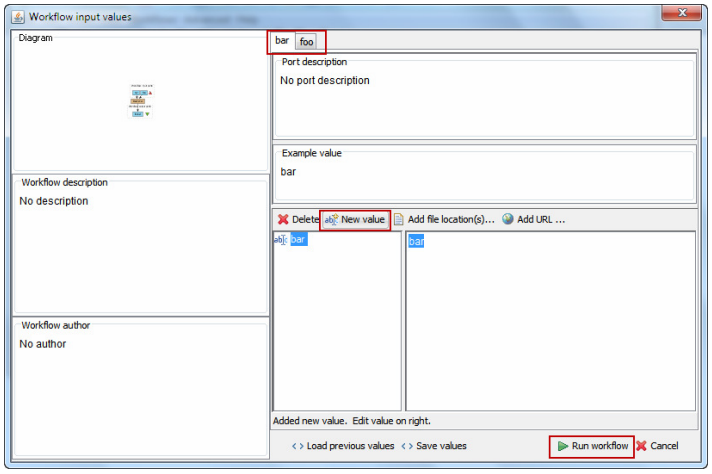
The workflow should be opened, and click the **CaGrid Service** icon to switch to the CaGrid Service perspective.



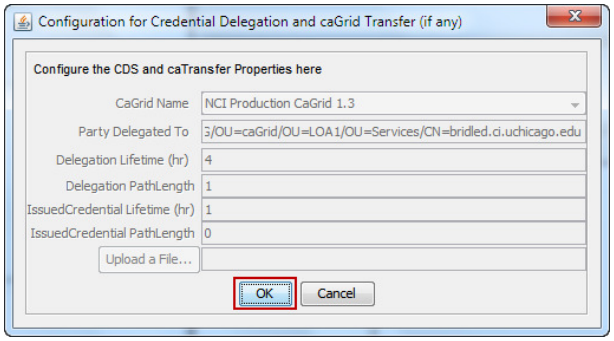
There is a working caGrid workflow service hosted at the University of Chicago, https://bridled.ci.uchicago.edu:5000/wsrf/services/cagrid/TavernaWorkflowService. You can change the URL to point to another service instance if you are sure it is working. Click **Run Workflow** button.



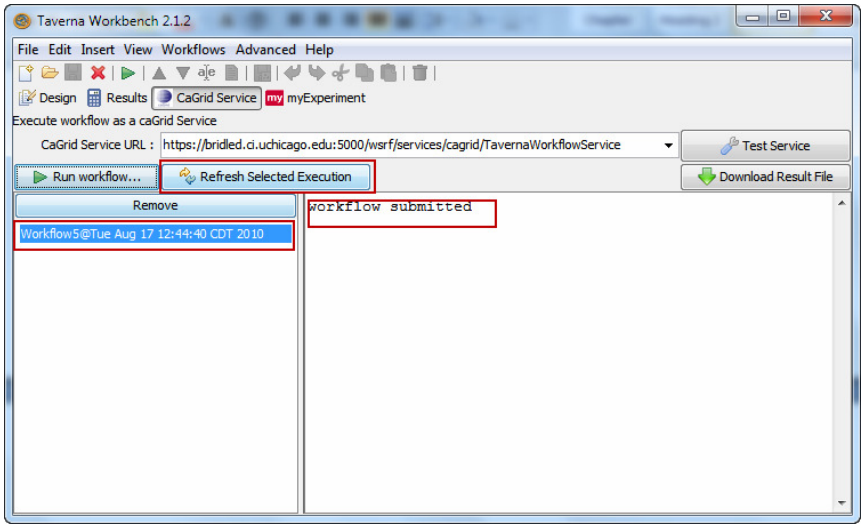
In the **Workflow input values** dialog, give a value to both inputs (i.e., *foo* and *bar*) in the tabbed view. For each tab, click **New Value** button and accept the default value provided (string “foo” and “bar”, respectively). Click **Run workflow**.



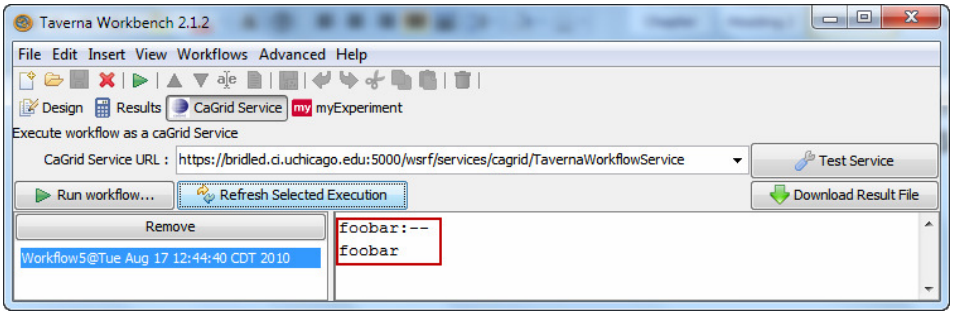
Click **OK** in the next dialog (we do not have to do anything related to CDS and transfer in this simple workflow).



Close the input values dialog and return to the workbench. The workflow has been submitted to the service. After a couple of seconds, click **Refresh Selected Execution** button.



You should be able to see string “foobar:--foobar” in the output pane, which means the workflow’s output port *foobar* has got a result value “foobar”. Congratulations, your installation has been validated.



4. World's Fair

4a. Posters

World's Fair – Exhibit Hall C

Poster Titles At-A-Glance

(The full abstracts can be found on pages 51-126. The category grouping and poster number reflects the location of the poster.)

Clinical Data and Trials Management

1. caBIG® Lab Viewer: Transforming Heterogeneous Data into Sharable, Standardized Information
2. caBIG®-compatible Software at Baylor College of Medicine's Dan L. Duncan Cancer Center
3. BRIDG: The International Standard for Protocol-Driven Research
4. Leveraging NCI Enterprise Services at DCP
5. Adopting and Piloting the caBIG® Adverse Event Reporting System (caAERS)
6. The caBIG® Adverse Event Reporting System (caAERS)
7. Using the PRO-CTCAE System to Solicit and Collect Patient Reported Outcomes
8. Open-Source, Standards-Based Decision Support Services: a Potential Enterprise Resource for caBIG®?
9. Evaluation of caBIG® Tools for the National Children's Study
10. The caBIG® Suite Implementation Update at the UAMS
11. A System for Indexing Clinical Documents for Clinical and Translational Research
12. Czech Clinical Registries Monitoring the Use of Targeted Drugs in Cancer Therapy
13. caEHR QA
14. caBIG® Integration Hub
15. Clinical Connector
16. Clinical Connector C3D: caGrid Service for C3D
17. CDA-based CRF Management Prototype
18. CDA-based CRF Management Architecture
19. caBIG® Deployment at Duke Comprehensive Cancer Center (DCCC)
20. Production Implementation of the Oncology Patient Enrollment Network (OPEN) - caBIG® Principles At Work

21. Achieving Continuity of Care and Enhancing Clinical Trials Success with the caBIG® Clinical Trials Suite
22. imHealthe™
23. NCI Enterprise Services (NES) for Structured Computable Representation of Protocols
24. Registering Trials with NCI's CTRP System
25. The Integrated Platform for Agents and Diseases (IPAD)
26. The caBIG® Central Clinical Participant Registry
27. C3PR NCI Enterprise Services Application in a Multi-site Environment
28. Supporting International Collaboration in Research Through the Use of Transferable, Clone-able Technology Solutions
29. Clinical Trials Reporting Program (CTRP)
30. CTMS for Traditional System of Medicine & Herbal Products
31. Advancing Medical Developments Through a Trusted Bridge: How the National Cancer Institute (NCI) Cancer Therapy Evaluation Program (CTEP), Pharmaceutical Company A, and SAFE-BioPharma Are Using Digital Identities to Improve Authentication and Eliminate Paper-based Forms

Imaging

32. Proposed Clinical Study Web Portal for the NHLBI Pediatric Heart Network – Using caBIG® and Open-source Technology for Image and Clinical Data Management
33. Tumor Imaging Metrics Core: A Centralized Service for Standardized Tumor Measurements for Multi-center Oncology Clinical Trials
34. Migrating National Lung Screening Trial's CT Image Library to a National Biomedical Imaging Archive
35. The caBIG® Annotation and Image Markup (AIM) 3.0 and AIM Template Creator
36. Adapting omniVisGrid Services for Mobile Histology Image Browser
37. Image Annotation Tool for Cancer Lesion Tracking and Automated Response Assessment
38. NBIA Implementation and Use Cases at Fox Chase Cancer Center
39. OmniSpect: A caBIG® Silver-level Certified Analytical Service for Unmixing Multispectral Images
40. National Biomedical Imaging Archive
41. NBIA Deployment for Adopters Including NCCCP Sites

Molecular Characterization

42. Implementation of a Systems Biology Data Integration Platform
43. Nanotechnology Characterization
44. nano-TAB: A Standard File Format For Data Submission and Exchange on Nanomaterials and Characterizations
45. Grid-Based Cancer Model Simulation with CViT's Computational Model Execution Framework
46. Identification of Dysregulated Networks in Acute Myeloid Leukaemia Using caBIG® Distance Weighted Discrimination
47. CNSuite: A caBIG® Analytical Tool for Copy Number Analysis
48. Leveraging Workflows in a Laboratory Information Management System
49. omniBiomarker 2.0: Extending a caBIG®-certified Application for Next-generation Sequencing
50. CaArray Standalone Client For Auto Importing and Uploading Array Files
51. geWorkbench: Offering Integrated Access to caBIG® and MAGNet Tools
52. caIntegrator2 - A Translational Research Tool for 21st Century Biomedicine
53. caNanoLab: Developing a Collaborative Environment Supporting the Application of Nanotechnology in Biomedicine
54. caBIO in GeneAnswers: Integrated Interpretation of Genes
55. Cancer Bench-to-Bedside (caB2B): Current State and Future Directions
56. DDN: A caBIG® Analytical Tool for Comparative Network Analysis

Biospecimen Collection

57. Management of Clinical Study Biospecimens by Integration of caTissue and ClinPortal, A Locally Developed Clinical Data Management System
58. High-throughput Biorepository Tools
59. Establishing the Connectivity Between Biospecimens and Clinical Data for Enabling Translational Research and Biomedical Discoveries
60. Tracking Biolgical and Environmental Samples for Large Population-Based Epidemology Studies
61. An Interoperable Framework to Link Biospecimen With Gene Array Data: A caArray and caTissue Integration Scenario
62. UCSF-RTOG Cooperative Group Biospecimen Banking

63. University of California San Francisco Helen Diller Family Comprehensive Cancer Center (UCSF-HDFCCC) Tissue Core Biospecimen Banking Effort

64. The GBC Reporting Tool: A Web-Based Catalog of NCI Cooperative Group Trial Biospecimens

65. Interoperable Biorepositories, Heterogeneous Systems, and Multi-Institutional Collaboration

66. Continued Development of caTIES, A Collaborative Tissue Banking and Text Mining Tool

67. Connecting TissueMetrix to the caBIG® Using the CBM Database and Web Service

68. Developing a Custom User Interface for Real-time Data Entry into caTissue

69. Adoption and Adaptation of caTissueSuite at the University of Iowa

70. Sharing Biospecimen Inventory From Locally-Developed Databases Using the CBM

71. Integration of caTissue with Clinical Data in an i2b2 Research Data Mart

72. Sharing Biorepository Information through the Common Biorepository Model (CBM) and NCI Specimen Resource Locator

73. Informatics as the Mechanism AND the Goal: How Interoperability Will Support the Cancer Human Biobank (caHUB)

Infrastructure

Vocabulary and Common Data Elements

74. Harmonizing the National Institute of Neurological Disorders and Stroke (NINDS) Core Common Data Elements (CDEs) with the National Cancer Institute (NCI) Cancer Biomedical Informatics Grid® (caBIG®) Cancer Data Standards Registry and Repository (caDSR)

75. Forms Curators - What They Do and How They Do It

76. Finding Information on the caDSR Content Wiki

77. Connecting EVS with the Semantic Web

78. Towards an Unambiguous and Formal Description of Cancer Therapy Experiments

79. Ontology-Based Queries for the caGrid Infrastructure

80. Curating All Data Elements In A Clinical Trial with CDEs from caDSR

81. Population-Based Data Collection Utilizing Open-Source Software and Common Data Elements

82. Tools Used to Create High Quality Common Data Elements (CDEs)

83. caBIG® and REDCap, CTSA and UCSF HDFCCC CRI - Incrementally Advancing Translational Informatics
84. Implementation of CTCAE Version 4.0 in C3D Studies
85. caDSR Bulk Loading Process
86. Safety Profiler
87. Ontology Modeling of caBIG® Semantics
88. The GeneTegra Information Integration System for caGrid Data Services
89. Planning and Execution of Queries against caGrid Data Services
90. Extending HL7-LOINC-NCI Terminologies for Reporting Genetic Test Results
91. LexEVS 6.0
92. Common Terminology Services 2 (CTS 2)
93. Harmonizing Clinical Study Data Elements in Heterogeneous Meta-Data Repositories Using the CSHARE/LexWiki Environment
94. LexWiki and Web-Protg Integration: A Comprehensive Framework of Collaborative Authoring for Large-scale Biomedical Terminologies
95. Grouping of caDSR Common Data Elements Using Non-metric Clustering of Concept Identifiers
96. Deployment of LexEVS for Enterprise Terminology Services: Lessons Learned
97. Collaborative Development of Ontologies using WebProtg and BioPortal
98. NCI EVS Terminology Browsers: Serving and Connecting a Growing Community
99. Sharing Behavioral Measure Data Using caGrid

Architecture

100. Life Sciences Subject Matter Experts Model Business Processes and Capabilities to Support ECCF
101. xService
102. Information Representation Working Group Information Model Development To Support ECCF
103. caAdapter Model Mapping Service: Tooling Support for ISO 21090 Development
104. caAdapter Data Mapping and Transformation Service: A Bridge to caBIG® Interoperability

105. LS DAM: A Foundation for Building Semantically Interoperable Services In The Life Sciences Domain
106. caGrid Portal
107. QA Initiatives at NCI CBIIT
108. Enterprise Service Specifications Team (ESST)
109. The Laboratory Information Digital Data Exchange (LIDDEx) Consortium: Two Years of Experience with the Development of a Semantically-encoded Interoperability Layer for Clinical and Investigative Laboratory Results
110. Leveraging caGrid Technologies for Linking Clinical and Biospecimen Data Repositories
111. Triton: An Integrative Translational Research Information Management Platform
112. Using the Open Metadata Registry (OpenMDR) to Generate Semantically Annotated Grid Services
113. Adaptation and Adoption of caBIG® Tools to a Clinical Research Community
114. caGrid 1.4
115. CTRP Trial Registration Service: NCI's ECCF Compliant Enterprise Service
116. caBIG® Interoperability Scenarios
117. The Cancer Therapy Evaluation Program Enterprise System (CTEP-ESYS) Service-Oriented Architecture (SOA)
118. Revisiting Address Interoperability in a SAIF Environment
119. NET-based Clients + Services + Cloud for Cancer Bioinformatics
120. ISO 21090 Adoption: The Plan and the Tools
121. Plugin Architecture for caBIG® Applications: Empowering the Adopter Community
122. Life Sciences Subject Matter Experts Model Business Processes and Capabilities to Support ECCF
123. Construction of a Standards-Based Platform-Independent Information Model for the Minnesota Congenital Heart Network
124. Web-based Semantically Integrated Framework for Cancer Research
125. Leveraging Cancer Bioinformatics Infrastructure Objects (caBIO) for Research and Experimental Annotations
126. caBIO Case Study: Development of an Enterprise Level Molecular Annotation Service Leveraging the NCI ECCF

127. Using GEM to Foster Sharing and Collaboration on the Grid
128. caLIMS2: Next Generation Cancer Laboratory Information Management System
129. The TCR Cancer Registry: A Case Study of caCore Based Data Standards Implementation to Integrate with the Cancer Biomedical Informatics Grid® Designed to be caBIG® Interoperable
130. caEHR Information Management - A Framework for Artifact Development and Deployment
131. Setting a New caBIG® "Openness Standard" - Building a Better "Network"
132. caGrid WebSSO
133. Common Security Module v4.2
134. CSM GAARDS Migration Module (CGMM) v0.6
135. Instance and Attribute Level Security: High Performance, Out-of-Box Solution From CSM
136. Service Development Tools for ISO 21090 Datatypes
137. Single Sign-On/Sign-Out for caGrid Enabled Web Applications
138. caCore Workbench
139. caCore Software Development Kit (SDK)

Training and Outreach

140. Accessing and Consuming caBIG® Training
141. caBIG® Learning Management System: A Tool for Getting Connected with caBIG®
142. caBIG® caGrid Knowledge Center
143. Ushering in Change, the caBIG® Way
144. Assembling Biomedical Collaborative Intelligence Using Google Wave Federation Protocol
145. VIVO and eagle-i: Why They are Important for caBIG®
146. caBIG® Vocabulary Knowledge Center
147. C-DAC: CBIIT Collaboration - Areas of Interest
148. User Interfaces: Designing Software to Look Like It REALLY Works Together

Poster Abstracts

1. caBIG® Lab Viewer: Transforming Heterogeneous Data into Sharable, Standardized Information

caBIG® connects the cancer research and clinical communities through open source, interoperable technologies that facilitate data sharing and collaboration. One successful example of caBIG® technologies in action is the caBIG® Clinical Trials Suite. The Suite is a collection of interoperable software tools that provide a comprehensive solution to managing clinical research and clinical trials information. Laboratory results are a key component of this clinical research information and play an important role in the clinical research lifecycle. Lab results make up a large percentage of clinical case report form data and may be used to calculate dose levels for the next trial phase or to determine if treatment outcomes are falling within the expected range. As the number of specialized laboratory and on-site/point-of-collection (POC) devices increases, so does the stream of laboratory data relevant to clinical trials. However, this data stream may originate from heterogeneous clinical laboratory systems or centralized labs that are not part of a research center's integrated clinical trials management system. What should be available electronically, in real time, must now be converted from the data structure of one system to another. These discontinuous procedures create discrepancies in data and between systems and hinder the overall clinical trial process. The solution is caBIG® Lab Viewer. Lab Viewer, through the use of caAdapter, provides a mechanism for converting heterogeneous laboratory data into usable clinical trials data. caBIG® Lab Viewer facilitates the automatic capture, translation, and import of data from clinical systems into caBIG®-compatible clinical trials databases by providing clinicians with the ability to view clinical laboratory data imported from clinical chemistry and other lab systems, then select and send that data to adverse event reporting or other clinical data systems via the Suite's caBIG® Integration Hub.

AUTHORS AND AFFILIATIONS: Jennifer Brush[1]; Naveen Amiruddin[1]; Becky Angeles[1]; Smita Hastak[1]. [1] ScenPro, Inc.

2. caBIG®-compatible Software at Baylor College of Medicine's Dan L. Duncan Cancer Center

The Dan L. Duncan Cancer Center (DLDCC) of Baylor College of Medicine has been a part of caBIG® since 2005. During that time, we have contributed to several caBIG® workspaces and had some significant successes deploying caBIG®-compatible software. These deployments were driven by real needs of the DLDCC; caBIG®-developed reference implementations were chosen after a market survey revealed they represented the most functional, cost-effective solutions. Production deployments include two instances of caTissue Suite 1.1.0 (caTS) and one instance of C3PR. The two instances of caTS consist of an internal instance supporting two virtual biorepositories and a public instance supporting a collaboration between the 11 prostate SPOREs. BCM will share a useful subset of data elements for >90,000 prostate biospecimens via the public instance, and other Prostate SPOREs will be sharing similar datasets. C3PR is used by DLDCC's Clinical Trials Support Unit (CTSU) to manage all therapeutic clinical trials' patient registration. We have slightly extended this instance of C3PR, identified four new data elements that will be added to C3PR 2.9, and developed a linked, partially caBIG®-compatible companion application to support CTSU needs that were deemed out-of-scope for C3PR. Currently, we are migrating 10 legacy breast cancer biobanks to the private caTS instance and working to implement Teleforms to convert and hand-written clinical form data to into

dynamic extensions, which we expect to dramatically improve the efficiency of data entry into caTS. In the future, we expect wider use of caTS to support several additional biorepositories, inclusion of non-therapeutic trials in C3PR, and expanded use of the tools in the Cancer Clinical Trials Suite, specifically PSC and caAERS. The resources we required to accomplish these goals and the lessons learned in the process will be discussed.

AUTHORS AND AFFILIATIONS: David Steffen, Ph.D. [1]; Pam Mayfield[1]; Jon Barney[1]; Lauren Boyd, Ph.D.[1] [1]The Dan L. Duncan Cancer Center, Baylor College of Medicine.

3. BRIDG: The International Standard for Protocol-Driven Research

The Biomedical Research Integrated Domain Group (BRIDG) Model is a collaborative effort of stakeholders from the Clinical Data Interchange Standards Consortium (CDISC), the HL7 Regulated Clinical Research Information Management (RCRIM) Work Group, the National Cancer Institute (NCI) and the US Food and Drug Administration. Their mission is to produce a shared view of dynamic and static semantics that collectively define the domain of clinical and pre-clinical protocol-driven research and its associated regulatory artifacts. BRIDG 3.0.2 (scheduled to be released summer 2010) will include harmonized semantics from the NCI's caBIG® Clinical Trials Suite 2.2. The Suite is comprised of a collection of the Patient Study Calendar (PSC), caBIG® Adverse Event Reporting System (caAERS), and LabViewer. BRIDG has been balloted and approved in the May 2010 ISO Joint Initiative Council (JIC), HL7, and CDISC standard ballot processes. Comments are being addressed and remediated - look for BRIDG to be an ISO, HL7, and CDISC standard in the near future. With the increase in the breadth of semantics in BRIDG and the number of BRIDG-based implementations, it has become difficult for domain experts to understand and recognize where their concepts map into the model and for technical experts to understand what parts to include in their implementations. As a result, the BRIDG Board of Directors (BoD) and the Semantic Coordination Committee (SCC) have redesigned BRIDG into multiple layers. The three layers allow stakeholders to view the content in terms they are familiar with. The upper layer is comprised of several UML-based subdomain views using domain-friendly terminology. The middle layer is a single UML-based view of the harmonized subdomain semantics along with and OWL-DL representation. The lower layer is comprised of several HL7 RIM-based models representing the harmonized domain semantics using non-ambiguous RIM concepts.

AUTHORS AND AFFILIATIONS: Becky Angeles[1]; Steve Sandberg[1]; Smita Hastak[1]; Wendy Ver Hoef[1];Julie Evans[2]; Mike Woodcock[3]; Lloyd McKenzie[4]; Charlie Mead[5] John Speakman[5]. [1]ScenPro, Inc.; [2]CDISC; [3]Mayo Clinic; [4]Gordon Point Informatics; [5]NCI CBIIT.

4. Leveraging NCI Enterprise Services at DCP

The NCI Division of Cancer Prevention (DCP) is one of the early adopters and consumers of NCI's Trial Registration service, an NCI Enterprise Service (NES). The objective of the new 73 DESK-CTRP Integration Module (DCIM) is to facilitate the submission of DCP-sponsored cancer clinical trials to the Clinical Trials Reporting Program (CTRP) System. The new module provides DCP with the capability of locating and submitting approved trials in the DCP Enterprise System Knowledgebase (DESK) to CTRP using the Trial Registration Service. It also provides a comprehensive audit trail to track trial submission activities across multiple platforms. Once trial information is transferred and abstracted by the Clinical Trials Reporting Office (CTRO), it will become available in CTRP, along with all

other NCI-sponsored cancer trials, to all authorized parties, which in turn facilitates knowledge exchange and helps support and advance cancer research in general.

AUTHORS AND AFFILIATIONS: Troy Budd[1]; Anne Tompkins[1]; Charles Yaghmour[2]; Hugh Reinhart[2]; Travis Riggs[3]; SAIC-F: Karen Parness[4]. [1]NCI DCP; [2] ScenPro, Inc.; [3]iFission; [4]SAIC-Frederick.

5. Adopting and Piloting the caBIG® Adverse Event Reporting System (caAERS)

The caBIG® Adverse Event Reporting System (caAERS) is currently being adopted by several institutions worldwide, including a cooperative group and several comprehensive cancer centers. Prior to adopting caAERS, each of these organizations has not only identified an unmet need for an adverse event system, but they have also determined their own internal readiness to successfully adopt and the readiness of caAERS to fulfill their requirements. Additionally, the adopting organizations and the caAERS development team must ensure adherence with the requirements of sponsors and regulatory authorities. A final, but critical component of system implementation is a validation of the system readiness to all stakeholders. The use of a controlled pilot study has been used to demonstrate the production readiness of caAERS at Mayo Clinic Rochester and piloting of caAERS at additional institutions is planned to begin shortly. Discussed further are the lessons learned from the initial pilot and several considerations that need to be made for future production usage of caAERS.

AUTHORS AND AFFILIATIONS: Paul Baumgartner[1]; Ram Chilukuri[1]; Sharon Elcombe[2]; Sonja Hamilton[2]; Kim Johnson[3]; Nimesh Patel[3]; Todd Crook[3]; Bob Morrell[3]; Ken Quinn[4]; Ann Setser[6]; Shanda Finnigan[6]; Mary Agnes Templeton[6]; Michael Ames[7]. [1]Semantic Bits, LLC; [2]Mayo Clinic; [3]Cancer and Leukemia Group B; [4] Wake Forest Health Sciences; [5] Roswell Park Cancer Institute; [6]NCI CTEP; [7]NCI CBIIT; [8] University of Colorado Cancer Center.

6. The caBIG® Adverse Event Reporting System (caAERS)

The caBIG® Adverse Event Reporting System (caAERS) is a standards-based, open-source software system that is used to record and report adverse events observed during clinical trials. Users can access caAERS through an intuitive Web 2.0 user interface or via several programmatic interfaces that support robust integration with multiple data sources including: clinical trial management systems, clinical data management systems, electronic health records, authentication / authorization systems, caBIG® Clinical Trials Suite applications, NCI's Enterprise Services, and various terminology systems including CTCAE and MedDRA. caAERS features a state-of-the-art rules engine, which provides a configurable system to automate the assessment and reporting of adverse events in compliance with regulatory and protocol requirements. The system also provides a powerful work-flow management feature, which facilitates routing adverse event reports to the appropriate personnel for review, comment, and approval before submission. Collectively, these features make caAERS a highly configurable and powerful adverse event reporting system that supports clinical trial business processes ranging from small scale, single site, stand-alone operations to large scale, multi-site, integrated enterprises.

AUTHORS AND AFFILIATIONS: Biju Joseph[1]; Srini Akkala[1]; Sameer Sawant[1]; Ion Olaru[1]; Monish Dombla[1]; Karthik Iyer[1]; Wesley Wiggins[1]; Denis Krylov[1]; Paul Baumgartner[1]; Vinay Kumar[1]; Ram Chilukuri[1]. [1]Semantic Bits, LLC.

7. Using the PRO-CTCAE System to Solicit and Collect Patient Reported Outcomes

The PRO-CTCAE software system is an open-source, standards-based, caBIG® compatible web application that provides tools for researchers and patients. The system is used to solicit and record

responses directly from patients regarding symptoms they have experienced during the course of a clinical trial. These patient reported symptoms (or outcomes) are based upon symptoms defined in the Patient Reported Outcome version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). This system supports authoring of PRO-CTCAE case report forms (CRFs), which are the questionnaires used to solicit patients' symptom experiences. This system collects patients' responses to these questionnaires via a simple and intuitive web-based interface. This system enables close monitoring of patient symptoms and possible adverse events occurring during cancer clinical trials. The system provides researchers with the ability to configure notifications regarding the status of CRF completion as well as enable alerts to clinical staff that are triggered based upon the characteristics of patient-reported adverse symptoms. Patient and study level reporting features are also included to enable review of symptom trends throughout the duration of a trial. The usability of the PRO-CTCAE system is currently the subject of a study at Duke Comprehensive Cancer Center and the feasibility of widespread use of the PRO-CTCAE system is the subject of a large multi-site study.

AUTHORS AND AFFILIATIONS: Harsh Agarwal[1]; Mehul Gulati[1]; Ram Chilukuri[1]; Vinay Kumar[1]; Paul Baumgartner[1]; Ethan Basch, M.D.,M.Sc.[2] [1]Semantic Bits, LLC; [2] Memorial Sloan-Kettering Cancer Center.

8. Open-Source, Standards-Based Decision Support Services: a Potential Enterprise Resource for caBIG®?

As caBIG® migrates towards a Services Aware Interoperability Framework (SAIF), a potentially key enterprise service that could be provided as a part of caBIG® is an open-source, standards-based clinical decision support service that supports various inferencing capabilities relevant to clinical research and personalized medicine. For example, a standards-based decision support service could help identify patients that appear to be eligible for specific clinical trials based on electronic health record data. Similarly, such a service could help provide patient-specific diagnostic and therapeutic recommendations based on clinical trial protocols. In particular, a promising resource that could be leveraged by caBIG® is a clinical decision support Web service developed at Duke University known as SEBASTIAN. SEBASTIAN is currently in widespread operational use at the Duke University Health System to support point-of-care chronic disease management and enterprise care quality reporting. Moreover, this service is being operationally used by large tertiary care hospital in Argentina and by North Carolina Medicaid, and it is in the process of being deployed at the Dana-Farber Cancer Institute in an NCI-funded study aimed at optimizing lung cancer symptom management. Beyond its operational validation, SEBASTIAN has served as the basis of the Health Level 7 (HL7) and Object Management Group (OMG) Decision Support Service standards, and the technology is currently being adapted into a fully open-source platform with no intellectual property restrictions. Moreover, the caBIG® Central Clinical Participant Registry (C3PR) team at Duke University is actively exploring how this service could be integrated within the C3PR architecture, in particular for identifying patients potentially eligible for recruitment into clinical trials. As caBIG® migrates to SAIF, this specific resource, or other open-source, standards-based decision support services, could potentially be incorporated into the caBIG® infrastructure and provide a core infrastructure capability supportive of caBIG®'s efforts to accelerate clinical research and enable personalized medicine.

AUTHORS AND AFFILIATIONS: Kensaku Kawamoto, M.D., Ph.D.[1]; Wesley Wiggins, M.S.[2]; Pankaj Agarwal, M.S.[3] [1]Center for Health Informatics, Duke University; [2]SemanticBits, LLC; [3]Comprehensive Cancer Center, Duke University.

9. Evaluation of caBIG® tools for the National Children's Study

We conducted a preliminary evaluation of the suitability of the caBIG® Suite and Biospecimen Management tools in combination with other open-source tools to the informatics requirements of the National Children's Study (NCS). The NCS aims to examine the effects of environmental influences on the health and development of 100,000 children across the nation. The study will collect data from the participants starting before birth until age 21. Our results thus far indicate that the functionality of the tools will meet the required operations and data-collection needs. One important limitation is the lack of functionality to manage the recruitment process prior to obtaining informed consent in the provider-based recruitment workflow. System Design and Implementation: We conducted a proof-of-concept analysis of our caBIG®-based clinical trials management suite (CTMS) housed at the UAMS. For this evaluation we used a test environment of the UAMS CTMS, which comprises the caBIG® C3PR, caBIG® PSC, caTissue, OpenClinica, and LimeSurvey. We created the study in C3PR and registered test subjects there. We created the NCS calendar in the PSC. The NCS calendar is complex, requiring different activities at different times for the mother, father, child, and other possible participants such as neighbors and other family members. We also have created planned biospecimen collection in caTissue. We created case-report forms in OpenClinica, an open-source, electronic data-collection tool. Conclusion: Our existing, open-source informatics infrastructure for conducting cancer trials at UAMS appears flexible enough to accommodate the vast majority of the needs of the NCS, a non-cancer study. The results show that our UAMS CTMS is promising to standardize our research data collection efforts in compliance with national initiatives. The NCS study center at our institution has decided to proceed with the UAMS CTMS, and to evaluate options for pre-consent recruitment tracking for inclusion into the suite.

AUTHORS AND AFFILIATIONS: Umit Topaloglu Ph.D.[2,3]; Cheryl Lane[2]; William R. Hogan M.D., M.S.[2,3]; Jiang Bian[2]; Thomas Kieber-Emmons Ph.D.[2]; Laura F. Hutchins M.D.[2]; Charlotte Hobbs, M.D,. Ph.D.[1] [1]Arkansas Children's Hospital Research Institute; [2]Winthrop P. Rockefeller Cancer Institute; [3]Division of Biomedical Informatics, University of Arkansas for Medical Sciences.

10. The caBIG® Suite Implementation Update at the UAMS

The caBIG® was launched to create a semantically interoperable informatics infrastructure to facilitate collaboration for cancer research. The Winthrop P. Rockefeller Cancer Institute at the University of Arkansas for Medical Sciences (UAMS) has decided to implement an open source and standards-based clinical research informatics framework employing the caBIG® Suite tools and its standards. After successful workflow development, the Cancer Institute has entered 132 open and active trials to the CTMS and 1407 subjects were registered to the same system. System Design and Implementation: UAMS Clinical Trials Management System (CTMS) was implemented as a portal with Single Sign On (SSO) that includes the following applications: caBIG® Central Clinical Participant Registry (C3PR), caBIG® Patient Study Calendar (PSC), OpenClinica (open source Clinical Data Management System from Akaza Research), UAMS Event Tracker (ET), UAMS TrialSearch, caBIG® caTISSUE, and caBIG® caARRAY. caBIG® caTIES was also implemented for de-identification and concept code generation of the Pathology Reports. In addition, caBIG® Labviewer (with auto CTCAE toxicity grading), caBIG® Adverse Event Reporting System (CAAERS), and PRO-CTCAE (the patient reported outcomes database) are being tested. For a successful adoption, the aforementioned open source tools were modified (as needed) and integrated with each other (e.g. OpenClinica caXchange integration) as

well as other enterprise systems such as ARIA (Grant and Protocol Submission and Tracking), and CRIMSON (Protocol Budgeting and Clinical Trial Contracting). Conclusion: Through this initiative, UAMS is leveraging valuable resources and saving time, which will allow Cancer Institute investigators to perform their research easier and collaborate with others with minimal effort. Our success has led several campus champions to become highly supportive of this informatics initiative and they have partnered with the team to roll this out for non-cancer research initiatives within the recently NIH/NCRR supported Center for Clinical and Translational Research and other UAMS initiatives.

AUTHORS AND AFFILIATIONS: Umit Topaloglu, Ph.D.[1]; Cheryl Lane[1]; William R. Hogan, M.D.[1]; Jiang Bian[1]; Tong Jett[1]; Ildeniz Yayla[1]; Zhidan Feng[1]; Thomas Kieber-Emmons, Ph.D.[1]; Laura F. Hutchins, M.D.[1] [1]Winthrop P. Rockefeller Cancer Institute at the University of Arkansas for Medical Sciences.

11. A System for Indexing Clinical Documents for Clinical and Translational Research

The vast majority of clinical data exist only as text. Unlocking the rich information within clinical documents for automated analysis therefore has the potential to enable new research discoveries and to facilitate translational science. We developed a system based on open-source tools, including Mirth Connect and caTIES, that indexes pathology and radiology reports generated during clinical care at UAMS Medical Center. We implemented a web-based client to query the documents. We have indexed over 200,000 documents to date. We are currently working through regulatory and compliance issues to make the system available to researchers. System Design and Implementation: As shown in the Figure, Mirth Connect, an open-source, standards-based healthcare integration engine, extracts, filters, and transforms HL7 messages from our registration and transcription systems into a centralized and structured data repository. Next, the caBIG® cancer Text Information Extraction System (caTIES) indexes free text reports including those from radiology and pathology, using various Natural Language Processing (NLP) techniques with controlled terminologies from the Unified Medical Language System (UMLS). The open source “Harvard Scrubber” bundled with caTIES provides de-identification functionality. Finally, we built a web-based client to allow researchers to query the newly unlocked data to identify patient cohorts based on various findings in the reports. Conclusion: We believe that this system based on open-source tools has great potential to extract useful biomedical information from unstructured text. Next steps include evaluation of the utility of the system.

AUTHORS AND AFFILIATIONS: Umit Topaloglu, Ph.D.[1]; Cheryl Lane[1]; William R. Hogan, M.D., M.S.[1]; Jiang Bian[1]; Vincent Yip[1]; Thomas Kieber-Emmons Ph.D.[1]; Laura F. Hutchins M.D.[1] [1]Winthrop P. Rockefeller Cancer Institute at the University of Arkansas for Medical Sciences.

12. Czech Clinical Registries Monitoring the Use of Targeted Drugs in Cancer Therapy

Apart from population-based data, the Czech Society for Oncology has been using a set of clinical registries to collect real clinical data on cancer patients in the form of medical documentation. The principal objectives of these projects involve monitoring and retrospective evaluation of the treatment results and safety in areas where treatment effectiveness and health care improvements are of particular importance. Special attention is paid to the modern treatment approaches based on new agents, which are specifically targeted at cancer cells. These clinical registries have been developed as non-interventional studies within the Czech National Cancer Control Programme (www.onconet.cz) and have been running in compliance with the Czech legislation in force, meeting all requirements with respect to security of the

collected data (for more details, see www.registry.cz). Primary objectives of clinical registries: (1) Monitoring the number of patients treated with new agents; (2) Assessing the treatment safety in terms of standardized toxicity scoring; and (3) Assessing the treatment effectiveness in terms of treatment response and survival. Secondary objectives of clinical registries: (1) Analysis of patients' survival in relation to the monitored clinical parameters; and (2) Analysis of data from patients treated with new agents in relation to the reference population data. The network of clinical registries is well organized and fully functional, covering all main Czech health care facilities dealing with cancer patients. Data collection is representative enough, bringing significant added value to both doctors and patients.

AUTHORS AND AFFILIATIONS: Rostislav Vyzula[1]; Ladislav Dusek, Ph.D.[2]; Petr Brabec[2]; Stepan Tucek, Ph.D.[1]; Alexandr Poprach[1] Daniel Klimes, Ph.D.[2 1] Masaryk Memorial Cancer Institute; [2] Institute of Biostatistic and Analyses Masaryk University.

13. caEHR QA

caEHR will develop and deploy business capability services that address the unique needs of the ambulatory oncology sector. The caEHR services include the standards-based specifications and the reference service implementation. These can be used by the implementers to build the oncology specific functionality in their Electronic Health Record (EHR) systems. The caEHR Quality Assurance (QA) team will validate the caEHR specifications and reference service implementations and provide User Acceptance Testing (UAT) support for NCCCP early adopter sites. The caEHR services must be Enterprise Compliance and Conformance Framework (ECCF) compliant to support interoperability, and must meet numerous regulatory requirements (HIPAA, Section 508, 21 CFR Part 11 and others). All artifacts developed throughout the project life cycle need to be validated for compliance against the relevant requirements and regulations. The caEHR QA team will use appropriate tools and review processes to ensure compliance. The caEHR QA strategy has been developed to encompass the project requirements and scope, while operating within the Agile methodology framework, and meeting the project timelines. Our strategy employs top-down system level testing focused on tests derived from Use Case, Architecture and Regulatory Compliance, and UAT requirements. Additionally we use bottom up service-level testing focused on tests derived from the Platform Independent Model (PIM) specifications to ensure all functional and non-functional requirements are validated. QA will employ Continuous Integration (CI) techniques, including a comprehensive integrated test automation strategy to realize a high level of automation and low-effort production of metrics to quantify quality. All QA artifacts will be published for public consumption.

AUTHORS AND AFFILIATIONS: Jeff Pierce[1]; Brian Bell[1]. [1]NCI CBIIT.

14. caBIG® Integration Hub

caBIG® Integration Hub (iHub) is a service-oriented, robust, configurable messaging infrastructure for exchanging information among various applications and systems. It provides support for reliable messaging, reliable transactions, transformations, email notifications, payload validation, audit logging, and high availability. iHub provides the infrastructure for integrating between grid, non-grid, multi-grid environments, and legacy applications. iHub supports integration with caTissue. The TRANslational Informatics System to Coordinate Emerging Biomarkers, Novel Agents and Clinical Data (TRANSCEND) project leverages this capability for integration between Tolven electronic Clinical Health Record (eCHR) system and caTissue. iHub is the integration engine for caBIG® Clinical Trials Suite (Suite). The

Suite leverages iHub for integration among the suite applications and integration between suite applications and CDMS such as C3D and OpenClinica. In addition, the Suite uses the iHub for integration with COPPA (Correlation, Organization, Person, Protocol Abstraction) Enterprise services. iHub's flexible architecture will enable quick integration with new Enterprise services as they become available. iHub is based on open source technologies (Apache Servicemix) and caBIG® infrastructure tools (Common Security Module (CSM), Common Logging Module (CLM), and Log Locator Tool (LLT). It leverages caGrid security infrastructure services, such as Credential Delegation Service (CDS), Dorian, and Authentication Service, to provide seamless integration and traceability across disparate applications.

AUTHORS AND AFFILIATIONS: Santosh Joshi[1]; Harsh Marwaha[1]; Ajay Nalamala[1]; Kalpesh Patel[1]; Christo Andonyadis[2]; John Speakman[2]. [1]Ekagra Software Technologies; [2]NCI CBIIT.

15. Clinical Connector

Clinical Connector (CC) provides a generic Biomedical Research Integrated Domain Group (BRIDG) v2.1-based interface for integration with a Clinical Data Management System (CDMS). The Clinical Connector is a web service-based implementation conforming to NCI's Enterprise Conformance and Compliance Framework (ECCF). The Clinical Connector includes capabilities for sending Subject Registration messages and Lab Test Result messages to the associated CDMS. These messages can come from either caBIG® Clinical Trials Suite applications like C3PR, LabViewer or other legacy applications. Clinical Connector service is supported by caBIG® Integration Hub (iHub), which includes components that provide out of the box integration with Clinical Connector. CDMS Vendors such as Medidata, OpenClinica, Velos, and others can implement the Clinical Connector service for integration with the caBIG® Clinical Trials Suite via iHub. In an upcoming release, Clinical Connector capabilities will be merged with those of the Cancer Center Clinical Database (C3D) Connector. This updated Clinical Connector will provide a common set of interfaces and capabilities for integrating with any CDMS.

AUTHORS AND AFFILIATIONS: Patrick Conrad[1]; Santosh Joshi[1]; Harsh Marwaha[1]; Christo Andonyadis[2]; John Speakman[2]. [1] Ekagra Software Technologies; [2]NCI CBIIT.

16. Clinical Connector C3D: caGrid Service for C3D

The C3D Connector provides a semantically integrated service layer via caGrid that allows C3D adopters to expose functions within C3D (Oracle Clinical CDMS). The exposed service layer uses a BRIDG based model and defines service operations that could be implemented by other CTMS systems. The poster will include a description of the Clinical Connector and its role in the CTMS Suite; detailed descriptions of each of the services provided by the connector; and the latest security features and processing enhancements to the data retrieval services.

AUTHORS AND AFFILIATIONS: Patrick Conrad[1]. Ekagra Software Technologies.

17. CDA-based CRF Management Prototype

Clinical Document Architecture (CDA) is a standards-based document exchange architecture provided by HL7. It provides a templated approach for creating structured (machine-readable and human-understandable) documents for exchange within or amongst enterprises. CDA-based CRF Management Prototype is an end-to-end implementation of CDA and demonstrates its use for the purpose of creating and managing Case Report Forms (CRFs) in a Clinical Trials environment. It comprises of the following three main modules. A Template Designer and

Publisher module allows the user to create templates for form fields (entry-level templates); modules (section-level templates) and finally the CRF itself (document-level templates). It also provides a mechanism to add various constraints on these templates and publish them as validation rules along with the templates. The Instance Editor module is a CRF data-entry module which uses the templates as input and renders a CRF allowing users to enter data. Once the form is populated by the user, the tool checks the data against the validation rules and saves the completed CRF data instance in the CDA format. Finally the Instance Viewer module that allows users to view or print the CRF data instances in human readable format. Additionally the data can be saved locally in CDA XML format or in CDISC's ODM format for transfer to a Clinical Data Management System (CDMS).

AUTHORS AND AFFILIATIONS: Amit Srivastava[1]; Kunal Modi[1]; Santosh Joshi[1]; Bob Dolin, Jingdong Li[2]; Christo Andonyadis[3]. [1]Ekagra Software Technologies; [2]Semantically Yours LLC; [3]NCI CBIIT.

18. CDA-based CRF Management Architecture

Clinical Document Architecture (CDA) is a standards-based document exchange architecture provided by HL7. It provides a templated approach for creating structured documents that can be exchanged across an enterprise. The proposed CDA-based CRF Management Architecture describes creating and managing reusable CRF Templates as well as validating and storing CRF Instances. It proposes two services: A CRF Template Management Service that aids in the creation and publication of CRF Templates and supporting artifacts (validation rules, implementation guides and entry forms); and a CRF Data Management Service that validates the CRF instances using the CRF Template's validation rules and then persists them in the CDA format for later use. It also demonstrates how the CRF Infrastructure leverages the Metadata Infrastructure to semantically annotate the entries (form fields) on a CRF, thereby making all of the data collected using these CRFs semantically interoperable. The architecture also showcases how any front-end application can leverage the CRF Infrastructure to render data-entry screens using the CRF Templates and capture data in a structured CDA format. This data can then be transformed and transmitted to back-end business services for persistence or further processing.

AUTHORS AND AFFILIATIONS: Kunal Modi[1]; Santosh Joshi[1]; Brian McIndoe[1]; Christo Andonyadis[2]; Dave Hau[2]. [1]Ekagra Software Technologies; [2]NCI CBIIT.

19. caBIG® Deployment at Duke Comprehensive Cancer Center (DCCC)

With multiple caBIG® workspaces and plenty of software tools to choose from, the DCCC took full advantage of the opportunity to pick the right tools to meet its needs. The diverse software solutions allowed us to meet program goals for the many groups and departments, filling software voids and supporting new directions in informatics. The Bio-Informatics shared resource is actively using caArray, a tool to guide the annotation and exchange of array data using a federated model of local installations. They are also using GenePattern, providing access to a broad spectrum of computational methodologies to analyze genomic data. In the Cancer Center Information Systems (CCIS) group, we have installed the cancer Adverse Event Reporting System (caAERS), used to collect and report clinical trial adverse events. CCIS also supports a locally hosted instance of the Cancer Central Clinical Database (C3D), an NCI adaptation of the Oracle Clinical database for collecting and reporting clinical trial data. In addition, the Cancer Central Clinical Participant Registry (C3PR) was implemented for registering patients, piloting the first multi-center clinical trial on this platform. caGrid is installed to enable us to share information and

analytical resources efficiently and securely, also allowing us to easily contribute to, and leverage the resources of other institutions. Ongoing development includes the adoption of caTissue, a biospecimen management tool designed for inventory control, tracking, and annotation. Future caBIG® adoption includes the Clinical Trial Suite (CTS), taking advantage of the utilities contained therein to facilitate interoperability among component resources. Our goal is to connect caTissue, CTS tools, and caArray through the caGrid or web services, creating a network within Duke helping physicians, nurses, statisticians, and researchers to efficiently share data internally and with the larger community. We also intend to integrate other commercial and locally developed software platforms with the caBIG® tools.

AUTHORS AND AFFILIATIONS: Robert Annechiarico[1]; Mohammad Farid[1]; Arati Nemani[1]. [1] Duke Comprehensive Cancer Center.

20. Production Implementation of the Oncology Patient Enrollment Network (OPEN) – caBIG® Principles At Work

In 2009, the Oncology Patient Enrollment Network (OPEN) was released by the Cancer Trials Support Unit (CTSU) as a web-portal for patient enrollments on to all NCI-sponsored Cooperative Group clinical trials. The success of OPEN is evidenced by the number of protocols (over 26) and the volume of enrollments (over 5,400), with weekly increase in adoption and utilization. The success of OPEN is due to the careful attention paid to operational support as well as the utilization of caBIG® principles and resources. OPEN utilizes the principles of semantic and syntactic interoperability, as it collects and distributes data from a federated network of data centers, each operating within different technology environments and employing the data in a meaningful and computable way for research purposes. OPEN leverages caBIG® tools such as the Form Builder, the caCore API, and the Common Security Model (CSM). CSM's user provisioning tool (UPT) was used to provide instance and attribute level security. OPEN uses standard webservice technology for data transport and the Clinical Data Interchange Standards Consortium's (CDISC) Operational Data Model (ODM) format for data interchange. The CTSU has adapted its existing customer support services for patient enrollment to support implementation and utilization of OPEN. These support services include development of eligibility checklists (ECs) within the OPEN environment, testing and release services for the electronic ECs, online training for sites users, Help Desk support, post-enrollment support for patient transfers, and project management support to coordinate implementation of protocols. This poster will demonstrate that a successful implementation hinges on both good technology and good support services.

AUTHORS AND AFFILIATIONS: Ravi Rajaram, M.S.[1]; Lucille Patrichuk, M.B.A.[1]; Jayan Nair, M.S.[1]; Kasi Perumal, B.E.[1]; Sushmita De, B.T.[1]; Diana Meyer, B.A.[1] [1]Westat.

21. Achieving Continuity of Care and Enhancing Clinical Trials Success with the caBIG® Clinical Trials Suite

Clinical trials are essential to the evolution of modern healthcare; however, they often present a segmentation of patient health information. The result of this segmentation presents patients and health care providers with obstacles to care delivery and discourages clinical trials enrollment subsequently hindering clinical research. The Center for Biomedical Informatics and Information Technology at the National Cancer Institute, has developed a comprehensive set of modular, interoperable, and standards-based tools, "the caBIG® Clinical Trials Suite," that bring technological solutions to solve this problem. The caBIG® Clinical Trials Suite facilitates data exchange and collaboration through efficient federation of data and the use of interoperable messaging and standards-based terminologies. Health care

information collected prior to diagnosis can be seamlessly retrieved by the clinical trial site at time of enrollment by using caBIG®'s interoperable clinical trial framework. Within the clinical trial setting caBIG® Clinical Trials Suite is used for enrollment, trial logistics, patient scheduling, recording, and reporting adverse events and incorporating other lab information to maximize treatment and trial success. Following the clinical trial, the patient's health record from both pre-clinical trial Electronic Medical Record (EMR), and the clinical trial record, can be retrieved and used to assure the continuity of care. Within the clinical trial framework, the ability to blend data from molecular medicine, treatment, pathology, and patient outcomes allows clinical researchers to develop and refine evidence-based strategies that are tailored to the individual's unique health needs. In summary, the implementation of caBIG® facilitates research and care through interoperability and connectivity across domains to achieve optimal outcomes for patients.

AUTHORS AND AFFILIATIONS: William T. Dyer Jr.[1]; John Speakman[2]; Todd Hardin[2]. [1] Pyramed Research; [2]NCI CBIIT.

22. imHealthe™

The imHealthe™ application, developed by CTIS, Inc., is an interoperable Personal Health Record (PHR) software solution currently available on both iPhone OS and Android OS platforms and is designed to overcome the many bottlenecks and redundancies existing in today's health care system that limit real-time access to individual patient health information. imHealthe™ digitally captures patient health information and allows real-time exchange between different web-based PHRs and health care providers and helps eliminate the need for redundant patient data entries and information assembling in single repositories resulting in improved efficiency and quality of patient care with reduced cost and errors. imHealthe™ combines the efficiency of a PHR (i.e., appointment scheduling, care management, reporting, billing, etc.) with the accessibility of computer-based records from multiple locations via a secure network. By using imHealthe™, providers can, regardless of where they are, readily access medical histories, lab test results, digitized diagnostic imagery and reference databases and overcome obstacles associated with globalization and adoption of PHRs in the health care domain. imHealthe™ is designed to seamlessly integrate with commercial PHR repositories, such as Google Health, Microsoft Health Vault, and Dossia and can interact with clinical research related repositories such as NCI's Patient Outcomes Services or other Hospital repositories. imHealthe™ supports patients by accurately documenting and communicating patient data and history to doctors and other healthcare providers via the convenience of mobile device and greatly reduces the potential for error or unrecorded data. imHealthe™ is easily adopted and adds value as a monitoring and tracking tool to provide the right information to the right person, anytime, in the palm of their hands.

AUTHORS AND AFFILIATIONS: Rajiv Shahi, M.Sc., P.M.P., MCAD[1]; Vivek Pillai, M.S., M.B.A., P.M.P.[1] [1]CTIS Inc.

23. NCI Enterprise Services (NES) for Structured Computable Representation of Protocols

The NCI's Clinical Trials Reporting Program (CTRP) is the first project at NCI to adopt the Services-Aware Interoperability Framework (SAIF) and design and implement a set of ECCF-compliant services, known as the COPPA services. These persons, organizations, protocols, and the correlations among them became the first foundational NCI Enterprise Services (NES). In support of the FDAAA law, the NCI has developed a core set of services and applications built on those services in order to support the registration and abstraction of trials. ScenPro developed

the core protocol services and apps to support the sites in registering their clinical trials with CTRP and ultimately with ClinicalTrials.gov. The trials registered with CTRP are abstracted using the Protocol Abstraction (PA) application that is built on COPPA services. The COPPA information model includes a BRIDG-based structured computable representation of protocols. These PA services of CTRP provide the first iteration of the structured computable protocol data and are currently limited to the requirements of registering a clinical trial to ClinicalTrials.gov. The structured protocol data focuses on planned study concepts for interventional studies - study definition, study objectives, outcome measures, and planned activities (substance administration, procedures, etc.) are some of the semantics provided by the PA services. As the CTRP project matures, these Protocol services within the NES stack will be extended to support Patient Accrual, Adverse Events, and Outcomes data in support of the CTRP program goals. With BRIDG as the Domain Analysis Model for the clinical research domain, these extensions to the structured protocol representation will continue to be BRIDG-aligned to achieve the ultimate goal of supporting working interoperability. These PA services are currently consumed by caBIG® Clinical Trial Suite as a source of record for all NCI clinical trials and provide a clear example of semantic-based working interoperability in the clinical domain.

AUTHORS AND AFFILIATIONS: Smita Hastak[1]; Naveen Amiruddin[1]; Hugh Reinhart[1]; Nellie Martchenko-Shimko[1]; Kalpana Guthikonda[1]; Anu Sharma[1]; Christo Andonyadis[2]; John Speakman[2]; Steve Silberman.[3] [1] ScenPro, Inc.; [2]NCI CBIIT; [3]SAIC-Frederick.

24. Registering Trials with NCI's CTRP System

Trial Registry is a web-based application for registering, updating, and amending of proprietary and non-proprietary interventional trials with NCI's Clinical Trials Reporting Program (CTRP) system. The purpose of the CTRP program is to establish a comprehensive database containing regularly updated information on all NCI-funded clinical trials. Grantees will enter a defined set of information for each clinical trial into the database through the CTRP Trial Registry system. Trial Registry supports various methods for submission, including User Interface (one trial at a time), and Batch Mode (multiple trials through an Excel-based upload mechanism). The Trial Registry application uses trial registration business service to submit the data to the CTRP system. Submitted protocols are abstracted and curated by the CTRP staff using the CTRP Protocol Abstraction System. Once trial abstraction is complete, the submitting PI will receive the abstracted data in a standardized XML format, which the PI can then submit to clinicaltrials.gov. The NCI will use the abstracted information to coordinate research efforts to optimize our nation's investment in cancer research. The information required to be submitted is aligned and harmonized with the clinicaltrials.gov reporting requirements. The CTRP Trial Registry system is built on the foundation provided by the NCI Enterprise COPPA Services. The COPPA services support the core concepts of Persons, Organizations, Protocols, and the Correlations.

AUTHORS AND AFFILIATIONS: Smita Hastak[1]; Nellie Martchenko-Shimko[1]; Naveen Amiruddin[1]; Hugh Reinhart[1]; Kalpana Guthikonda[1]; Anu Sharma[1]; Christo Andonyadis[2]; John Speakman[2]; Steve Silberman[3]. [1]ScenPro, Inc.; [2]NCI CBIIT; [3]SAIC-Frederick.

25. The Integrated Platform for Agents and Diseases (IPAD)

For over a decade, the Cancer Therapy Evaluation Program Enterprise System (CTEP-ESYS), designed to enhance the scientific and administrative aspects of cancer clinical trial development and management, has collected and maintained clinical data on protocol and patient information. The Integrated Platform for

Agents and Diseases (IPAD) improves the data search capabilities of the CTEP-ESYS and assists NCI personnel in meeting data analysis, data mining, and transactional computing needs by providing flexible query, analytical, and integrated data warehouse capability. It is comprised of an enterprise search engine built on a suite of scalable components designed to aggregate information across various data sources such as the CTEP-ESYS, Enterprise Vocabulary Services (EVS), the Division of Cancer Treatment and Diagnosis (DCTD) programs (e.g., Developmental Therapeutics Program (DTP), Cancer Imaging Program (CIP), etc.), PubMed, ClinicalTrials.gov, etc. The application integrates with NCI, Clinical Trials Reporting Program (CTRP) to utilize the data maintained by CTRP, such as study registration, study accrual, etc. IPAD is a unified query tool which queries both structured and unstructured data stored as network files, biomedical articles and life science journal abstracts from the PubMed repository, and study data located in registries from NCI/CTEP, pharmaceutical companies, cooperative groups, cancer centers, consortia and academia involved in the development of new and novel therapies to treat cancer. Users of IPAD have the ability to drill down results that are returned from queries to granular details and provide reports for tracking the timeliness of protocols in support of Operational Efficiency Working Group (OEWG) recommendations. IPAD includes a customizable user interface to enable users to set preferences and generate reports and outputs to aid in their analysis and to export those outputs in various formats (e.g., PDF, Word, PowerPoint, etc.).

AUTHORS AND AFFILIATIONS: Steven Friedman, M.H.S.A.[1]; George Redmond, C.G.E.I.T., M.Sc., M.B.A.[1]; Shanda Finnigan, R.N., B.S.N., C.C.R.C.[1]; Sudhir Raju, M.B.A., M.S., P.M.P.[2]; Amit Vengurlekar, M.S., P.M.P.[2]; Sanjeev Kakar, B.E.[2]; Yvette Grainger, M.S.[2] [1]NCI, DCTD, CTEP; [2] CTIS, Inc.

26. The caBIG® Central Clinical Participant Registry

The caBIG® Central Clinical Participant Registry (C3PR) is a web-based application used for end-to-end registration of patients to clinical trials. This includes capturing data related to informed consent, eligibility, stratification, randomization, and screening. Clinical trial workflows are enabled by subject and study-site centric views into the registration process. Registration details are captured for all types of studies including companion studies, such as quality of life and tissue banking studies. The C3PR data model is harmonized with the BRIDG model version 3.0.1 and the application is compatible with the caBIG® Silver Level Compliance criteria. C3PR can be run in a standalone mode where study definitions, investigators, study personnel, and sites are entered into the system or in an integrated mode with the caBIG® Clinical Trials Suite (The Suite). C3PR integrates with the NCI Enterprise Services (NES) to ensure that the latest persons, organizations, and studies are present in the system and displayed to the user. All data sharing within C3PR is enabled through caGrid services and leverages standard grid security protocols for user management, authentication, authorization, and trust. C3PR also exposes the core functionality through services based on the Enterprise Conformance and Compliance Framework, which include Subject Management, Subject Registration, and Randomization services. The services will also enable multi-site clinical trials using installations of C3PR at two or more sites. Throughout the development of C3PR, a number of elaborator and adopter sites are actively being engaged to help define requirements and test the application. Our primary elaborators include Duke, Wake Forest, Georgetown, Mayo, Westat, CALGB, CCR, and the Coalition of Cooperative Groups. Our primary adopters include Duke, Wake Forest, Georgetown, and University of Arkansas for Medical Sciences with

engagement of Mayo and Baylor College of Medicine.

AUTHORS AND AFFILIATIONS: Pankaj Agarwal, M.S.[1]; Ram Chilukuri, M.S.[2]; Wes Wiggins, M.S.[2]; Vijaya Chadaram, B.S.N.[1]; Robert Morrell, M.B.A.[3]; Sharon Elcombe, M.A.[4]; Steve Riordan, M.B.A.[5]; Jieping Li, M.B.A.[6]; Robert Annechiarico[1]. [1] Comprehensive Cancer Center, Duke University; [2]Semantic Bits; [3] Comprehensive Cancer Center, Wake Forest University; [4]Mayo Clinic; [5]Westat, Inc.; [6] Lombardi Comprehensive Cancer Center, Georgetown University.

27. C3PR NCI Enterprise Services Application in a Multi-site Environment

The caBIG® Central Clinical Participant Registry (C3PR) is a web-based application used for end-to-end registration of patients to clinical trials. This includes capturing the consent signed date, eligibility criteria, stratification, randomization, and screening. Clinical workflows are enabled by subject-/study-/site- centric views into the registration process. The primary capabilities of C3PR are management of subject information and registration of subjects to clinical trial studies. These capabilities are now available as NCI Enterprise Services, developed within the framework of the Service Aware Interoperability Framework's (SAIF) Enterprise Conformance and Compliance Framework (ECCF). These services can be used by any external application, within or outside the caBIG® program, for interoperating with C3PR to achieve management of subject information and subject registration. C3PR can be deployed in three different modes: standalone, hosted, and multi-site. The primary goals of C3PR multi-site functionality are to facilitate clinical trials workflows between a coordinating center and the affiliate sites while reducing double data entry and manual transmission of data. Multi-site registrations can be done through the C3PR NES. Study data can be transmitted programmatically from the coordinating site to the participating sites where additional information, such as the IRB approval date and study personnel, is entered, and the study is opened locally for accrual. Subject Registration details are entered locally at the participating site and transmitted automatically to the coordinating site where enrollment and randomization is performed. These services are deployed at each site with operations for subject registration. Data transmission is secured using the ws-security infrastructure, which also leverages the industry standard PKI technologies. Business rules and notifications are enforced both locally at participating sites as well as at the coordinating site, allowing for robust data capture while maintaining data integrity.

AUTHORS AND AFFILIATIONS: Pankaj Agarwal, M.S.[1]; Ram Chilukuri, M.S.[2]; Wes Wiggins, M.S.[2]; Vijaya Chadaram, B.S.N.[1]; Robert Morrell, M.B.A.[3]; Sharon Elcombe, M.A.[4]; Steve Riordan, M.B.A.[5]; Jieping Li, M.B.A.[6]; Robert Annechiarico[1]. [1]Comprehensive Cancer Center, Duke University; [2] Semantic Bits; [3]Comprehensive Cancer Center, Wake Forest University; [4] Mayo Clinic; [5] Westat Inc.; [6]Lombardi Comprehensive Cancer Center, Georgetown University.

28. Supporting International Collaboration in Research Through the Use of Transferable, Clone-able Technology Solutions

NIAID is supporting projects to curtail infectious diseases that primarily affect specific regions of the world (e.g., malaria, leishmaniasis, schistosomiasis, and Ebola), as well as diseases that have the potential to cause global pandemics (e.g., influenza, SARS). As part of its global research agenda, NIAID is working to develop prevention and therapeutic strategies adapted for the unique needs of developing countries; build and sustain research capacity in-country; stimulate scientific collaboration and global partnerships; and work with in-country scientists to develop training, communications, and outreach programs. NIAID recognizes that a solid foundation of international research and collaborations enhances the U.S. capacity for infectious disease surveillance and the

ability to respond to newly emerging disease threats. NIAID pursues strategic methods to develop and evaluate relevant international infrastructures and collaborations with governments and communities that will multiply the effective use of resources and facilitate host organizations in building up their research programs. The Institute has utilized OpenClinica for several prototype studies conducted internationally that will be duplicated and propagated to multiple additional sites by the host nation or organization. The open-source CDMS tool is uniquely situated to be cloned as a technology solution, allowing it to transfer freely when propagated studies emanate from the initial project that are not Institute-funded or managed. The tool's flexibility and viability enable the transfer of the full investigative solution spectrum to the collaboration partner.

AUTHORS AND AFFILIATIONS: John McCarthy[1]; Cynthia Lassnoff[1]; Bianca Koscielski[1]; Nitesh Patel[1]; Jorge Monltalvo[1]; Chris Whalen[1]; Joe Croghan[1]. [1] NIAID Office of Cyber Infrastructure and Computational Biology.

29. Clinical Trials Reporting Program (CTRP)

One of the recommendations of the 2005 the Clinical Trials Working Group Report was the creation of a comprehensive database of information on all NCI funded clinical trials to facilitate clinical trial planning and prioritization. In response, the National Cancer Institute (NCI) launched the Clinical Trials Reporting Program (CTRP) in January 2009. To date, more than 2000 interventional clinical trials have been registered by NCI grantees. The system will support the reporting of patient accrual data in late 2010. Future enhancements include the capture of adverse event and outcome data, as well as the production of draft reports that contain data used in the preparation of Summary 4 reports. The Clinical Trials Reporting Program will utilize NCI's Enterprise Services (NES). The evolution of NCI's information products will include the reporting of capture and storage of data in one virtual location, eliminating duplicate reporting for NCI-grantees. The NES model is supporting current efforts to integrate Person and Organization information from the Cancer Therapy Evaluation Program's (CTEP) Enterprise Core Module (ECM) to provide a single source of data for all NCI users.

AUTHORS AND AFFILIATIONS: Kimberly A. Eckley, M.A., P.M.P.[1]; Gisele Sarosy, M.D.[2]; Joseph Martucci, M.S.[3] [1]Lockheed Martin Corporation; [2] Kelly Services; [3]Essex Management.

30. CTMS for Traditional System of Medicine & Herbal Products

Though the Clinical Trial Management System (CTMS) available on caBIG® addresses the needs of modern medicine, it lacks the capability to deal with Traditional Systems (THM) and alternate medicine (CAM). THM/CAM require more extensive data management, hence a requirement for a customized CTMS. Modern drugs use a single synthetic chemical entity whereas THM uses complex natural organic molecules in their formulations. Further, the use of a holistic paradigm in THM emphasizes on the intrinsic ability of the human body to restore to the healthy state and hence emphasizes in great detail on the individual profiles of the patient such as: Patient constitution & persona details (prakruti, panchmahabhuta); defining the patient's disease type (Dosha); patient's clinical examination (Rogipareeksha); and parameters relating to patient's lifestyle and food habits. The core concept of Health and Disease in Ayurveda, a biochemical, bioenergetic and biospiritual system of medicine is built around the uniqueness of the individual using a threefold criteria known as the Tridosha theory which identifies principles of motion (VATA), metabolism (PITTA), and structure (KAPHA) as discrete phenotypic properties, i.e., PRAKRUTI based on differences in physical, physiological and psychological

characteristics. Traditional Chinese medicine (TCM) is also based on similar concepts. An enhanced and adopted conventional CTMS is thus required to deal with these systems. CTMS-TRA is an endeavour to bridge this gap. Herbal Medicine, though new to the western world, has been the preferred healing system in Asia, Africa, and Latin America for centuries. In many parts of the eastern world THM is the standard and western medicine is the alternative remedy. The poster presentation will showcase the concepts used to develop CTMS-TRA.

AUTHORS AND AFFILIATIONS: Madhulika Tripathi[1]; Ravindar Khanna[1]; Dr. Ranjit Prasad[1]. [1]First eLearning Quest Private Limited.

31. Advancing Medical Developments Through a Trusted Bridge: How the National Cancer Institute (NCI) Cancer Therapy Evaluation Program (CTEP), Pharmaceutical Company A and SAFE-BioPharma Are Using Digital Identities to Improve Authentication and Eliminate Paper-based Forms

The proposed abstract will explore how Pharmaceutical Company A, the National Cancer Institute's Cancer Therapy Evaluation Program (CTEP), and SAFE-BioPharma are collaborating to demonstrate how an all electronic workflow – facilitated by digital signatures – speeds up business process flow and improves the NCI's ability to accelerate research and be more responsive to public health. Pharmaceutical Company A directly and indirectly communicates with many organizations in NCI. Current methods include Federated Identify Management. Federal Bridge PKI methods further enhance the NCI's capabilities. Integrating these technologies with existing processes and using digital signatures for approvals will improve overall business processes. This will reduce the time required for trials, improve turn-around times on drug and protocol approvals, and quicken the time to market. Digital signatures will be used for review, approval, and signing of documents such as Letter of Intent, Concept Approval, Protocol Approval, Clinical Trial Agreements, and contracts. The pilot, which at this writing is underway, is expected to provide a model for broader use of Federal Bridge and SAFE-BioPharma compliant digital signatures in business process flows within NCI and NIH. The pilot has three primary components: (1) Cross certified credentials – Three different types of cross certified credentials from three different groups (Pharmaceutical Company A staff; NCI personnel and Cooperative Groups) that will be used for digitally signing documents and authentication; (2) Types of Documents – Documents being signed include, but are not limited to, Letter of Intent, Concept Approval, Protocol Approval, Clinical Trial Agreements and Contracts; and (3) Workflow – The pilot is configured to alert the signatory that a document needs to be signed and to alert the document owner when all signatures have been obtained. Because the Pharmaceutical Company A /NCI digital signing pilot eliminates any need for wet signatures – and therefore, any need for paper – it improves the speed and efficiency by which medical research is conducted. To put this in context, the New England Journal of Medicine has estimated that 40 percent of the cost of bringing a new drug to market is tied to paper-based processes.

AUTHORS AND AFFILIATIONS: Tanya Newton[1]; Cindy Cullen[1]. [1]SAFE BioPharma Association.

32. Proposed Clinical Study Web Portal for the NHLBI Pediatric Heart Network Using caBIG® and Open-source Technology for Image and Clinical Data Management

The National Heart, Lung, and Blood Institute's Pediatric Heart Network (PHN) started in 2001 as a collaboration of medical centers to conduct nationwide research studies in children with heart disease. We have worked with the PHN to identify their

requirements for a web-portal based on open-source tools. This portal may be used for their image-based, multi-center clinical trials to track the status of image submissions from clinical sites to the data coordinating center and to core labs, and provide access to analysis results. One critical need is to provide a way to track a center's submitted images through quality control steps and data processing. Identification and assessment of caBIG® and other technologies was conducted, and resulted in a prototype web portal. The PHN Portal is a web interface to access remotely located image archive(s) and centrally-stored associated clinical data for use in reading and annotating the images. The Portal was developed using open-source tools, including the RSNA Clinical Trials Processor (CTP), as well as components of the caBIG® National Biomedical Imaging Archive (NBIA). These tools enable the transfer of a clinical image study from one hospital PACS system, through appropriate security, to the remote image archive; and can incorporate anonymization and de-identification. The initial demonstration tested the integration of these tools/archives to a useful, workflow-tracking portal. The portal demonstrates the ability to track data submission status via a "dashboard", perform validation and quality control of images and their clinical data, and download images for reading. Additional features can include study management, viewing of submitted cardiac ultrasound echo image studies (single image and CINE loop viewer), and PHN role-specific access and views. The prototype version of the PHN Portal showcases how open-source image archive and web tools are used to provide a clinical trial tracking system to allow participating centers to view the status of submissions; synergies with CVRG and using caGrid are also under consideration. This is supported with funds from NHLBI-NIH Grant No. U01HL068270 for support of the Pediatric Heart Network.

AUTHORS AND AFFILIATIONS: A. T. Fernandez, Ph.D.[1]; D. C. Wu[1]; T Shi[1]; K. Johnson[1]; M. A. Keller, Ph.D.[1]; S. Tennstedt, Ph.D. [2]; G. Pearson, M.D.[3]

Booz Allen Hamilton; [2]New England Research Institute; [3]NIH National Heart, Lung, and Blood Institute.

33. Tumor Imaging Metrics Core: A Centralized Service for Standardized Tumor Measurements for Multi-center Oncology Clinical Trials

The Tumor Imaging Metrics Core (TIMC) is a Dana-Farber/Harvard Cancer Center (DF/HCC) imaging core laboratory which provides tumor measurements of radiological scans for oncology clinical trials. The TIMC performs standardized measurements (RECIST, WHO, Cheson, SUV, 3D volume, etc.) for CT, MR, and PET imaging studies according to trial protocol. Target and non-target lesions are tracked longitudinally and results are stored in the TIMC database on a secure website accessible by authorized trial staff. Scan analyses are ordered and results are viewable online.The DF/HCC established the TIMC in order to deliver standardized tumor measurement services to clinical trials in an efficient and cost-effective manner. The TIMC is a multi-institutional entity with laboratory space at three locations: Massachusetts General Hospital, Brigham and Women's Hospital, and Dana-Farber Cancer Insitute. TIMC services are available to other Cancer Centers through secure Internet and wide-area network connections. Imaging studies are accessible from the caBIG® grid or other secure medical image network services providers. The TIMC provides four primary services for clinical trials: 1) Consultation for protocol and image analysis design (as requested); 2) Image capture of radiological scans from local or remote sites; 3) Quantitative assessment of lesions from patient scans according to trial protocol; and 4) Secure web services, including online order-entry and multi-media results reporting. The DF/HCC Tumor Imaging Metrics Core addresses an outstanding need for efficient and consistent delivery of standardized radiological measurements for multi-center

clinical trials. The TIMC has transformed the way in which tumor assessments are obtained at the DF/HCC. With its unique blend of expertise, advanced technology and quality service, TIMC has become strategically indispensable to the Cancer Center and serves as a model for the development of similar cores. For sites with limited imaging core laboratory services availability locally, TIMC services can readily accessed remotely using modern electronic pathways.

AUTHORS AND AFFILIATIONS: William Hanlon, M.S.[1]; Gordon Harris, Ph.D.[1]; Trinity Urban M.S.[1]; Marta Braschi, M.D.[1]; Vadim Frenkel, M.S.[1]; Edna Chiu, B.S.[1]; Anand Singh, M.D.[1]; Jamie Barerro, M.D.[1]; Vikram Kurra, M.D.[1] Gisele Cruz, M.D.[1]; Annick D. Van den Abbeele, M.D.[1] [1]Dana-Farber/Harvard Cancer Center, Massachusetts General Hospital, Brigham and Women's Hospital.

34. Migrating National Lung Screening Trial's CT Image Library to a National Biomedical Imaging Archive

From 2002-2004, the Lung Screening Study of the National Lung Screening Trial (NLST) enrolled 34,614 participants, aged 55-74, at increased risk for lung cancer due to heavy cigarette smoking. Participants, randomized to X-ray or CT arms, received up to three imaging screens for lung cancer at annual intervals. Screening is complete; final collection of participant follow-up data through 2009 is underway; an announcement of NLST primary endpoint results is anticipated in 2011. From 2005-2007, available CT exams (48,547) were uniformly de-identified and delivered to the CT Image Library (CTIL) at Washington University. Access to CTIL is currently limited to research projects approved by NLST. We plan to migrate CTIL to updated storage controlled by a local BlueArc network storage system. As part of that migration, we will change CTIL's management software and have considered the National Biomedical Imaging Archive (NBIA) software as one alternative. By so doing, CTIL could be prepared for public access following the announcement of NLST results: 1) To evaluate NBIA, we installed several instances of the NBIA software into an array of virtual and physical machines. In some cases, we installed older versions of the NBIA software, then upgraded to newer versions, allowing us to understand the process of successfully upgrading without a loss of stored data. We experimented with changes to the NBIA internal database schemas and query forms to better support the needs of NLST researchers. We created improved mechanisms for migrating existing repositories into NBIA, including mapping the CTIL into an NBIA-format, allowing proper de-identifying and re-identifying of images per NLST protocols. Migration and performance testing is now underway; and 2) We shall report on migration progress, database schema remapping, performance measures moving CTIL image studies from the old data repository to the new, and how the CTIL could be made available through the NBIA using caGrid-Services access.

AUTHORS AND AFFILIATIONS: David Maffitt[1]; Paul Koppel[1]; Stephen Moore[1]; Kenneth Clark[1]; David Gierada[1]; Lawrence Tarbox[1]; Fred Prior[1]. [1]Washington University.

35. The caBIG® Annotation and Image Markup (AIM) 3.0 and AIM Template Creator

The Annotation and Image Markup (AIM) standard has demonstrated its usefulness by integrating the descriptive information of an image with user-generated graphical symbols and textual descriptors placed on the image into a single common information source. The annotated information includes anatomic entity, characteristic of anatomic entity, imaging observation, characteristic of imaging observation and inference. The AIM data model collects many types of image semantic content using standard vocabularies such as RadLex, SNOMED CT, and DICOM, and customizable user-defined terminology. Existing vocabularies contain thousands of terms that make it difficult for multiple users to consistently

search and select terms for inclusion as an AIM annotation. In the majority of cases, the user desires to make simple and constrained annotations that is reproducible.An AIM template XML schema and web-based application for creation of an AIM XML document template have been developed to allow an expert to create an XML document containing controlled terminologies. It can be consumed by an application capable of understanding AIM and AIM template(s). A template allows users to rapidly choose appropriate semantic image content in a standardized format.The current AIM data model is version 3. It has undergone multiple revisions, public comments, and real world usage. The model now captures anatomic entity characteristic, inference, annotation role, AIM status as well as characteristic quantification. An inference provides a conclusion derived by interpreting an imaging study and/or medical history. Annotation role describes the role of referenced annotation. AIM status capture a status of an annotation instance using coded term, a version of annotation instance and update authorization. A quantification can be a numerical value, an interval (e.g. 34-67%), a scale (e.g. 1:None, 2:Mild), a quantile (e.g. 1(1-50), 2(51-100)) and a non-quantifiable (e.g., none, mild, mark). Thus, AIM 3.0 has evolved into a highly capable model and the template schema supplements its capabilities.

AUTHORS AND AFFILIATIONS: Pattanasak Mongkolwat, Ph.D.[1]; David S. Channin, M.D.[2]; Vladimir Kleper[1]; Vivek DaveM.D.[1]; Danial L. Rubin, M.D.[3] [1] Northwestern University; [2] Guthrie Healthcare System; [3] Stanford University.

36. Adapting omniVisGrid Services for Mobile Histology Image Browser

TissueWiki Mobile is an Apple iPhone application capable of browsing millions of histology images provided to the public by the Human Protein Atlas. These images display protein expression by histological staining using over 6000 antibody targets, and cancerous tissue images can be compared to normal tissue images of the same type to determine if staining is different. These images can validate cancer biomarkers, but the pathologist annotations of the images are brief and qualitative. We have developed automated processes that produce color segmentation masks and image quality scores and place all of this data into a publicly-available wiki at http://tissuewiki.bme.gatech.edu. We hope that allowing mobile device owners to assist in quantifying features of the images will result in an augmented data resource of use to cancer researchers. omniVisGrid services (currently under Silver Compatibility Review) were used to generate color histograms of the images in the RGB and HSV space. Users may annotate the images using the iPhone touch interface to allow for Region of Interest (ROI) selection or to assist in cell counting. Future work might include development of a cancer target library for new nanoparticle stains using vocabulary standards, or adoption of a microscopic imaging standard developed by other caBIG® Working Groups.

AUTHORS AND AFFILIATIONS: Chih-wen Cheng, M.S.[1]; Todd H. Stokes, Ph.D.[1]; Sovandy Hang[1]; Yachna Sharma, M.S.[1]; May D. Wang, Ph.D.[1,1] Emory University and Georgia Institute of Technology.

37. Image Annotation Tool for Cancer Lesion Tracking and Automated Response Assessment

Background: Radiological imaging provides rich information for evaluating the response of cancer patients to treatment. However, consistent acquisition of cancer lesion measurements over serial imaging studies and application of response criteria such as RECIST is time-consuming and error-prone. We sought to develop an open source tool to assist lesion tracking and automate assessment of tumor burden. Methods: We previously built a semantically-aware image annotation tool called iPAD, an open source plugin to the Osirix image viewing workstation. iPAD extracts and saves

quantitative and semantic image metadata in the caBIG® AIM standard format. We extended iPAD with analytic routines to process each annotated lesion, classify it as target or non-target lesions, and calculate quantitative features needed to assess the RECIST response criteria. Results: We evaluated our tool by using it to annotate lesions seen in serial imaging studies from patients in a cancer research study. The tool automatically generated target lesion treatment response assessments (categories of response) directly from the image annotations. An oncologist validated the treatment response assessments automatically derived from our tool. An evaluation of the tool for use in assessing treatment response is currently underway for a multi-site clinical trial. Conclusions: We developed and evaluated an image annotation tool to assist reviewers in tracking cancer lesions over time and to automatically apply RECIST response criteria directly from image annotations, eliminating the need to perform calculations and apply criteria by hand. The advantage of our tool is that all the information needed to derive automated assessment of tumor burden is directly derived from image annotations. This approach may improve the ability of oncologists and radiologists to use quantitative information in images to evaluate tumor burden.

AUTHORS AND AFFILIATIONS: Mia A Levy, M.D.[1]; Daniel L. Rubin, M.D., M.S.[2 1] Vanderbilt University Medical Center, Department of Biomedical Informatics and Medicine, Vanderbilt Ingram Cancer Center; [2]Stanford University School of Medicine, Department of Radiology and Biomedical Informatics.

38. NBIA Implementation and Use Cases at Fox Chase Cancer Center

The American College of Radiology (ACR) guidelines recommend radiologists read at least 50 endoscopically confirmed Computed Tomography Colonoscopy (CTC) (Virtual Colonoscopy or VC) cases to prepare for CTC interpretation. Ideally this collection of training cases will be chosen to demonstrate the gamut of appearances of colonic polyps and CTC interpretation pitfalls. Additionally, the cases should include examinations performed for a variety of indications and with varied acquisitions. Fox Chase has performed over 100 CTC exams, but requires access to a larger number of studies for training reasons. The NCI VC collection includes 808 cases that could be used for training purposes. Fox Chase would like to contribute to and update the NCI VC collection, with scans utilizing a 64 Slice Siemens CT and with newer imaging protocols to allow other institutions access to recent VC studies for training and research. To accomplish these training criteria we have chosen to implement NBIA 4.4.1 on a Linux-based server. Using the DICOM transfer protocol the images are sent from our Philips iSite PACS (3.6.64.0) installation to the Clinical Trial Processor (CTP) software developed by Dr. John Perry of RSNA for anonymization and submission into NBIA. After QC, the images are then made publicly available to all other NBIA instances, including CBIIT's via caGrid (www.imaging.nci.nih.gov). We anticipate submitting approximately 100 CTC studies per year, which will be made publicly available via caGrid. Future use of NBIA would include Fox Chase fostering increased use of NBIA by developing new and novel collections of training and research material based of Fox Chase teaching files. There is also interest in utilizing NBIA and its associated technologies to allow protocol management companies secure access to imaging studies for clinical trial patients. This poster will also present the implementation story as well lessons learned and possible workflow considerations for successfully implementing NBIA in your organization.

AUTHORS AND AFFILIATIONS: Brian Vecchiarelli[1]; Kristen Edwards, M.D.[1]; Michael Collins, M.S.[1]; Brian J. Hughes[2 1]Fox Chase Cancer Center; [2]Terrapin Systems.

39. OmniSpect: A caBIG® Silver-level Certified Analytical Service for Unmixing Multispectral Images

A large variety of imaging modalities collect multispectral data. Common color images capture three different wavelengths (red, green, and blue), whereas fluorescence and mass spectrometry images contain tens to thousands of single channel images. The spectrum at each pixel is commonly modeled as a mixture of underlying source spectra. For example, fluorescence images of quantum dot stained tissue samples contain a combination of the characteristic quantum dot spectra. Mass spectrometry images contain a combination of the characteristic tissue spectra. We present omniSpect, a caBIG® certified analytical service that unmixes multispectral images using reference spectra for known constituents.

AUTHORS AND AFFILIATIONS: R. Mitchell Parry, Ph.D.[1]; Todd H. Stokes, Ph.D.[1]; Sovandy Hang, B.S.[1]; Martin Ahrens, B.S.[2]; May D. Wang, Ph.D.[1]
[1]Department of Biomedical Engineering, Georgia Institute of Technology and Emory University; [2] College of Computing, Georgia Institute of Technology.

40. National Biomedical Imaging Archive

As imaging becomes an increasingly important part of cancer research, there is a greater need for integrative imaging platforms that support the integration of clinical, imaging, genomic, and proteomic data. The National Biomedical Imaging Archive (NBIA) is an initiative within the National Cancer Institute to provide a repository and web-based application, specifically providing support for image and annotation submission, retrieval, and download. The National Biomedical Imaging Archive (NBIA) offers the ability to remotely search and download metadata and DICOM images to be used in the development and validation of analytical software such as computer-aided diagnosis (CAD) tools through the use of caGrid technology, the underlying service oriented infrastructure supporting the cancer Biomedical Informatics Grid® (caBIG®). This service provides access to public images, including those from the TCGA collection (The Cancer Genome Atlas), RIDER collection (Reference Image Database to Evaluate Response), CT Colonography collection, etc. NBIA has been investigated and/or adapted by multiple research institutes and cancer centers worldwide through caBIG® program, such as the U.K. National Cancer Research Initiative (NCRI), Shanghai Center for Bioinformatics Technology (SCBIT), Pediatric Brain Tumor Consortium (PBTC), Memorial Sloan-Kettering Cancer Center (MSKCC), University of California San Francisco (UCSF), etc. Our Lady of the Lake Medical Center will host the NBIA to share image data with its sister site, the Mary Bird Perkins Cancer Center.

AUTHORS AND AFFILIATIONS: Robert Shirley[4]; Ed Helton[4]; Eric Kascic[1]; Qinyan Pan[1]; Jim Zhou[1]; Quy Phung[3]; Jill Hadfield[4]; John Freyman[2]; Justin Kirby[2]; Brian Hughes[8]; Eliot Siegel[5,6,7]. [1]Science Applications International Corporation (SAIC); [2]SAIC-Frederick, Inc.; [3]ESAC, Inc.; [4]NCI CBIIT; [5]University of Maryland Department of Diagnostic Radiology; [6]VA Maryland Healthcare System; [7]MIRC Committee, Radiological Society of North America; [8]TerpSys.

41. NBIA Deployment for Adopters Including NCCCP Sites

The National Biomedical Imaging Archive (NBIA) is a searchable repository of *in vivo* radiological images. NBIA provides cancer care centers, the cancer research community, and academia access to images in DICOM format with rich metadata, annotations, and markup. SAIC–Frederick, Inc. (SAIC-F) in association with Software Consultants, Inc. (SCI) have been funded by the National Cancer Institute (NCI) Center for Biomedical Informatics and Information Technology (CBIIT) to provide cancer centers, particularly through the National Cancer Institute Community Cancer Centers Program (NCCCP), with customized deployment support at no cost

to the centers for NBIA and supporting tools. We have refined a process for ensuring smooth deployments of NBIA to NCI-selected cancer research centers and community care centers. The deployment process begins with a kick-off meeting followed by a **site customization process.** Site customization begins by having the deployment site complete a Site Assessment Document detailing the deployment environment. The completed document guides interviews between SCI and deployment site SMEs and users. We identify data sharing and patient info de-identification needs. Proper de-identification is vital if information is to be shared. The Clinical Trials Processor (CTP) tool allows detailed custom de-identification policies and comes with HIPAA templates. Usually, two repositories of data are maintained: de-identified public research data and full information private clinical data. CTP allows a single PACS data feed to be de-identified as needed to create these repositories. Final customization steps involve writing use-cases, training, hardware recommendations, the deployment plan, and lastly, simulating the deployment environment in SCI's laboratory to validate the deployment plan. Deployment is installation and data migration, end-to-end testing, training, and user acceptance. Post deployment includes transition to NBIA help desk support, documentation revision, process improvement, and user feedback.

AUTHORS AND AFFILIATIONS: Anand Basu[1]; Sudha Chudamani[2]; Peter Yan[2]; Scott Halpine[2]; Ashley Jacobs[2]. [1]NCI CBIIT; [2]SAIC-Frederick.

42. Implementation of a Systems Biology Data Integration Platform

To enable and support a systems biology approach to research, one requires an underlying infrastructure to manage, integrate, and share high-throughput functional genomics data and workflows from data production through annotation, analysis, and knowledge acquisition. At its core, there should be a comprehensive data management and annotation system and data repository that fully support publicly established standards for storing and reporting high-throughput functional genomics investigations. Such a system will serve as the platform's central hub and it will integrate with data analysis, visualization, and mining tools. It will also enable collaboration between internal teams, publishing of internal investigations and data to public repositories, and incorporation of public investigations and data into the platform for internal comparison and analysis. Here, the implementation of a systems biology data integration and knowledge management platform to support experimental and computational workflows, examining *in vivo* and *in vitro* generated systems response profiles (gene expression, microRNA, comparative genomic hybridization, and reverse-phase protein array proteomics data) is reported. The platform utilizes open-source, freely available components where suitable, featuring caArray and caGrid from the National Cancer Institute Biomedical Informatics Grid (NCI caBIG®) software family as its core data management and annotation infrastructure. For data exchange, the community standard MAGE-TAB format was used. caArray is integrated with GenePattern, an open-source bioinformatics workflow management system for integrative genomics, which is used for quality check, data analysis, and visualization purposes. caArray is also integrated with several commercial data analysis and biological pathway inferencing systems. Gene-centric, cross-investigation data mining capabilities are provided by the BioMart and InterMine open-source data warehouse systems. Under development are several other modules to integrate the platform with existing laboratory information management systems, as well as additional features to contribute the open-source caArray project.

AUTHORS AND AFFILIATIONS: Leandro Hermida[1]; Sam Ansari, Ph.D.[2]; Svenja Diehl, Ph.D.[2]; Carine Poussin, Ph.D.[2]; Alain Sewer, Ph.D.[2]; Yang Xiang, Ph.D.[2]; Filipe Bonjour[2]; Komanda Phanzu[2]; Sylvain

Gubian[2]; Bruce O'Neel[2]; Julia Hoeng, Ph.D.[2]; Manuel Peitsch, Ph.D.[2] [1]Leandro Hermida Consulting; [2]Philip Morris International R&D.

43. Nanotechnology Characterization

The caBIG® Integrative Cancer Research Nanomedicine Working Group is currently reviewing the common characterization methods used in the nanomedicine domain. The purpose of this review is to identify the common informatics needs of these methods as well as to prioritize assays to be described by the new nano-TAB standard (see accompanying poster). Our current review has included assays from the National Cancer Institute's Nanotechnology Characterization Laboratory (caNanoLab), the Oregon Nanoscience and Microtechnologies Institute, the National Institute for Environmental Heath Sciences' National Toxicology Program, the Food and Drug Administration, and other related efforts. By using nano-TAB to provide a standard specification for widely-used characterization assays, the Nanomedicine Working Group hopes to: drive the submission and exchange of nanomaterials to/from nanotechnology resources like the NCI's caNanoLab nanotechnology portal and the Oregon State Nanomaterial Biological Interactions (NBI) knowledgebase; empower organizations to adopt standards for representing data in nanotechnology publications; and provide researchers with guidelines for representing nanomaterials and characterizations to achieve cross material comparison.

AUTHORS AND AFFILIATIONS: Nathan A. Baker[1]; Raul Cachau[2]; Brad Collins[3]; Gilberto Fragoso[4]; Elaine Freund[5]; Marty Fritts[2]; Sharon Gaheen[2]; Elizabeth Hahn-Dantona[4]; Stacey Harper[6]; Mark D. Hoover[7]; Fred Klaessig[8]; Juli Klemm[4]; Subhas Malghan[9]; David Paik[10]; Sue Pan[2]; Rebecca Reznik-Zellen[11]; Grace Stafford[12]; Todd Stokes[13]; Dennis Thomas[14]. [1]Pacific Northwest National Laboratory; [2]SAIC-Frederick; [3]National Institute of Environmental Health Sciences; [4]NCI CBIIT; [5]3rd Millennium; [6]Oregon State University; [7]National Institute for Occupational Safety and Health; [8]Pennsylvania Bio Nano Systems, LLC; [9]Food and Drug Administration; [10]Stanford University; [11]National Nanomanufacturing Network; [12]The Jackson Laboratory; [13]Georgia Institute of Technology; [14]Washington University.

44. nano-TAB: A Standard File Format For Data Submission and Exchange on Nanomaterials and Characterizations

The field of nanomedicine faces many challenges in the development of standards to support meaningful data submission and information exchange. Numerous physico-chemical, *in vitro*, and *in vivo* assays must be addressed, with measurements currently dependent on non-standardized protocols and diverse technology types. Representing Structure-Activity-Relationships in nanomedicine, in particular, is critical to understanding the effects of nanomaterial structure on biological activity. Unfortunately, information describing the nanomaterial including functionalizing entities and 3D structure is often represented in an undisciplined fashion. This lack of standardization has been a significant deterrent to meaningful data sharing across the nanotechnology community; few publications contain sufficient information to enable adequate interpretation of results and successful achievement of experimental reproducibility. The nano-TAB effort aims to address data sharing challenges in nanotechnology by providing a standard means for identifying nanomaterials and characterizations in a tab-delimited format. nano-TAB is based on existing standards developed by the European Bioinformatics Institute (EBI) and the Investigation/Study/Assay (ISA-TAB) file format, which represents a variety of assays and technology types. The nano-TAB specification leverages ISA-TAB files describing investigations, study-samples, and assays and provides extensions to support nanomaterial structural information and concepts from the Washington University NanoParticle Ontology (NPO). The nano-TAB standard specification will enable the submission and exchange of nanomaterials to/from nanotechnology resources like the NCI's caNanoLab

nanotechnology portal and the Oregon State Nanomaterial-Biological Interactions (NBI) knowledgebase; empower organizations to adopt standards for representing data in nanotechnology publications; and provide researchers with guidelines for representing nanomaterials and characterizations to achieve cross-material comparison. The nano-TAB effort is a collaboration between a variety of organizations including the NCI, Washington University, Oregon State University-ONAMI, NIOSH, Stanford University, and ISA-TAB. nano-TAB is registered as an International ASTM Work Item, which facilitates a broad community outreach and input to the development of nano-TAB and other standards needed to support nanomedicine.

AUTHORS AND AFFILIATIONS: Juli Klemm, Ph.D.[1]; Nathan Baker, Ph.D.[2]; Dennis Thomas, Ph.D.[3]; Stacey Harper, Ph.D.[4]; Mark D. Hoover, Ph.D.[5]; Marty Fritts[6]; Raul Cachau[6]; Sharon Gaheen, M.B.A.[7]; Sue Pan[7]; Grace Stafford, Ph.D.[8]; David Paik, Ph.D.[9]; Brad Collins[10]; Elaine Freund.[11] [1]NCI CBIIT; [2]Pacific Northwest National Laboratory; [3]Washington University; [4]Oregon State University, the Oregon Nanoscience and Microtechnologies Institute; [5]National Institute for Occupational Safety and Health; [6]SAIC-Frederick Nanotechnology Characterization Laboratory; [7]SAIC; [8]Jackson Laboratories;[9] Stanford University; [10]National Institute of Environmental Health and Safety; [11]3rd Millenium.

45. Grid-Based Cancer Model Simulation with CViT's Computational Model Execution Framework

The NIH/NCI-supported Center for the Development of a Virtual Tumor, CViT (PI: T. S. Deisboeck), brings together a multi-institutional, interdisciplinary group of investigators with interest in the biomedical, computational, and mathematical aspects of cancer research. To foster the collection and sharing of in silico cancer models, simulation-related workflow designs, and access and integration of relevant tumor biology data from disparate sources, CViT's Digital Model Repository (DMR) was implemented as a semantically-enabled web-based data store. The CViT DMR was expanded in 2008 to provide a caBIG® silver-level compliant data service in order to allow client applications to securely upload and access models and model metadata within the repository and is currently being made interoperable with an upcoming clinical cancer modeling repository in Europe with support by the European Commission. The Computational Model Execution Framework (CMEF) was developed in 2010 to enable the grid-based execution of the computational cancer models deposited within the DMR. Utilizing CMEF, members of the CViT community are able to select and configure models to be executed, determine the data to be used to run these models, and deposit simulation results back to the repository.The CMEF integrates seamlessly into the CViT.org website providing grid-based model execution of Java, C/C++, and R programs on 32- and 64-bit Windows and Linux nodes. This expansion of CViT has been performed without affecting the caBIG® silver-level compatibility of the CViT DMR data service. In addition, the functionality added to support the execution of models, and especially the semantic metadata added to the system, has been created complying with caBIG® silver-level compatibility guidelines in preparation of releasing the CMEF as a silver-level compliant analytical service. This poster describes the expansion of CViT's Digital Model Repository to support the Computational Model Execution Framework.

AUTHORS AND AFFILIATIONS: Thomas J. Taylor, M.S.[1]; Ray M. Bradley, B.Sc.[1]; Peter Vazquez, B.Sc.[1]; E. Patrick Shironoshita, M.S.[1]; Mansur R. Kabuka, Ph.D.[1]; Jonathan Sagotsky, M.A.[2]; Zhihui Wang, Ph.D.[2]; Thomas S. Deisboeck, M.D.[2] [1]INFOTECH Soft, Inc.; [2]Complex Biosystems Modeling Laboratory, Harvard-MIT (HST) Athinoula A. Martinos Center, Massachusetts General Hospital-East.

46. Identification of Dysregulated Networks in Acute Myeloid Leukaemia Using caBIG® Distance Weighted Discrimination

Acute Myeloid Leukaemia (AML) represents one of the most genetically heterogeneous malignancies, however, some processes seem to be commonly dysregulated. Here we present a strategy for identifying important gene networks that are dysregulated in AML where the data was derived from different sources and different array platforms. We used Affymetrix GeneChip® 3' expression microarrays to determine gene expression profiles (GEP) in human AML patients. GEP data were generated from AML patients enrolled in two different AML NCRI-MRC UK clinical trials using two different Affymetrix platforms, HG-U133A (n=216) and HG-U133Plus2.0 (n=139). GEP from normal individuals were downloaded from ArrayExpress (n=26). In order to compare AML vs. normal GEP and to increase the power of analysis, it was important to combine the GEP into a single dataset. Individual .CEL files were imported into Partek® Genomics SuiteTM and GC-RMA normalisation was applied. Mixed model analysis of variance (ANOVA) demonstrated that the different types of platform contributed a significant amount of data variation, thereby providing systematic bias. To remove the bias, we performed batch adjustment by using Distance Weighted Discrimination (DWD) method from normalized microarray data. Similar steps were also successfully adapted to MAS5.0 normalized data. DWD is a robust statistical tool for batch correction available in Java version that merged data based on similar genes shared by both platforms. The merged datasets showed significant reduction in the source of data bias with GEP clustered according to their biological variation rather than technical variation. Subsequently, we performed inferential statistical tests and threshold analysis (1.5 fold) from the adjusted data followed by gene enrichment analysis using MetaCore (GeneGo Inc.) to identify dysregulated genes in AML. In conclusion, we have shown that DWD adjusted data combined from multiple existing datasets allows the identification of differentially affected networks in a genetically complex disease such as AML.

AUTHORS AND AFFILIATIONS: Siti Sarah Daud, M.Med.Sc.[1]; Alan K. Burnett, M.D.[1]; Richard L. Darley, Ph.D.[1]; Alex Tonks, Ph.D.[1] [1]Department of Medical Genetics, Haematology and Pathology, School of Medicine, Cardiff University.

47. CNSuite: A caBIG® Analytical Tool for Copy Number Analysis

CNSuite (Copy Number Suite) is a caBIG® (cancer Biomedical Informatics Grid®) analytical tool for gene copy number change analysis. CNSuite consists of a Fused Margin Regression (FMR) method for detecting copy number changes in a single signal profile and consensus copy number changes in population data, and two feature indexing methods for analyzing chromosomal instabilities (CIN). FMR uses the first-order variable fusion constraint to enforce a piecewise constant profile of estimated copy numbers and the epsilon-insensitive loss function to measure the errors of using the estimates to approximate the observed copy number signals. Compared with model-based detection methods such as Hidden Markov Models, FMR has better sensitivity and specificity in detecting copy number changes in noisy signal profiles with complex copy number patterns. We implement a fast solution-path algorithm to solve the optimization problem associated with FMR, which makes FMR applicable to high-density microarray data. Based on the detection results of FMR, we propose a quantitative CIN measure to summarize the trend of copy number alterations in chromosomes. We also implement a Haar wavelets-based method for multi-resolution analysis of chromosomal instabilities. CNSuite is applicable to analyzing germline copy number variations (CNV) in the study of population genetics

and somatic copy number alterations (CNA) in tumor genomics. CNSuite is implemented in R and C++ and can be used on various operating systems.

AUTHORS AND AFFILIATIONS: Yuanjian Feng[1]; Lei Song[1]; Huai Li[2]; Ie-Ming Shih[3]; G. Steven Bova,[3]; Jianfeng Xu[4]; Subha Madhavan[5]; Robert Clarke[5]; Yue Wang[1]. [1]Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University; [2]Bioinformatics Unit, RRB, National Institute on Aging; [3]Johns Hopkins University School of Medicine; [4]Wake Forest University School of Medicine; [5]Lombardi Comprehensive Cancer Center and Department of Oncology, Georgetown University.

48. Leveraging Workflows in a Laboratory Information Management System

caLIMS2 is a Laboratory Information Management System (LIMS) currently under development that will promote an accurate flow of sample and associated experimental data to and through a laboratory resulting in information that can be used for subsequent scientific analyses. Many LIMS systems in use today utilize static workflows that often require recoding of software to allow for modification of processes. This is a limitation for LIMS which has a need for frequent process changes. The caLIMS2 application will contain an adaptive workflow process that will allow laboratory personnel to author, share, and execute workflows across laboratories. The caLIMS2 workflow module is a runtime environment that provides the capability to execute business processes. It essentially provides three types of services covering process definition, process execution, and administration of processes. By leveraging caBIG® workflows and services, the caLIMS2 workflow module will provide the ability to define common specialized laboratory activities, dynamic processes, and execute those processes. Workflow design is a multistep process that produces workflow model instances (descriptions), which are then published for execution to the workflow engine. Workflows provides the capability to capture laboratory provenance on data related to equipment, equipment parameters, authors, reagents, environmental conditions, samples, procedures, etc. Workflows can be used at both research facilities and high throughput core facilities. The caLIMS2 application will contain preloaded workflows of established standard laboratory operating procedures that laboratories can use right out of the box. By implementing an adaptive workflow caLIMS2 will be usable across a wide array of laboratory domains.

AUTHORS AND AFFILIATIONS: Michael Visee, M.S.[1]; Michael Connolly, M.S.[1]; Christophe Ludet, M.S.[1] [1]Moxie Informatics.

49. omniBiomarker 2.0: Extending a caBIG®-Certified Application for Next-Generation Sequencing

Next-generation sequencing (NGS) is an important emerging technology in high-throughput biomedical data analysis. Effective analysis of NGS data, as with microarray data, requires standardized methods for data sharing and integration. We describe an extension to omniBiomarker, a caBIG® Silver-certified analytical service, that enables analysis of NGS data as well as microarray data. The extended features of omniBiomarker include: (1) new models that describe NGS data; (2) analytical services for sequence alignment; and (3) improvements to the original omniBiomarker analytical grid services for biomarker identification from RNA-Seq data. These improvements to omniBiomarker are natural extensions of the analytical grid service that reflect the changing landscape of cancer bioinformatics.

AUTHORS AND AFFILIATIONS: John H. Phan, Ph.D.[1]; Chang F. Quo[2]; Jenny Dale[2]; Jaydeep Srimani[2]; Todd H. Stokes, Ph.D.[1]; R. Mitchell Parry, Ph.D.[1]; May D. Wang, Ph.D.[1] [1]Georgia Tech/Emory University; [2]Georgia Tech.

50. CaArray Standalone Client For Auto Importing and Uploading Array Files

caBIG® caArray is an open-source, web and programmatically accessible array data management system developed to guide the annotation and exchange of array data. It requires the array files to be uploaded manually using a web user interface. The number of the files generated in a genomics core and required time to upload all these files may jeopardize the extensive use of caArray. To automate the process of file upload and file import, we developed a standalone client application with a graphical user interface (GUI) for CaArray 2.2.1 System Design and Implementation. caArray Standalone Client (CSC) is developed using Enterprise Java Beans (EJB) technology to allow array lab data to be uploaded as a batch process to the caArray 2.2.1 web application. It requires the following user parameters to automatically upload and import: 1) target caArray address and user credentials; 2) Experiment name to upload the files into. If no experiment name is provided, it can generate an experiment based in a predefined way; 3) provider name to be picked from the target caArray; 4) array design files that explains the design of an array; 5) assay type that is defined by the target experiment; 6) source directory where the generated files are located; and 7) a batch processing option that allows the process to be performed one time or in given intervals (6 hrs, 24 hrs, etc.). We have also developed an ANT script for CSC to make it an easily deployable application. Conclusion: We built the CSC for automatic remote upload and import of caArray files into the web application. As a result of the work, lab personnel will save time that is required to upload files. We have made the CSC open source and publicly available to expand caArray usage in the community.

AUTHORS AND AFFILIATIONS: Umit Topaloglu, Ph.D.[1]; Sree Rama Konduri[1]; Cheryl Lane[1]; William R. Hogan M.D.[1]; Jiang Bian[1]; Laura F. Hutchins, M.D.[1]; Clyde Bailey[1]; John Shaughnessy Jr., Ph.D.[1]; Thomas Kieber-Emmons, Ph.D.[1] [1]Winthrop P. Rockefeller Cancer Institute at the University of Arkansas for Medical Sciences.

51. geWorkbench: Offering Integrated Access to caBIG® and MAGNet Tools

geWorkbench (www.geworkbench.org), a platform for integrated genomics analysis, brings together tools developed under the auspices of both the caBIG® program and the National Centers for Biomedical Computing (NCBC), as well as from other third-party sources. Through the NCBC-supported Center for the Multiscale Analysis of Genomic and Cellular Networks (MAGNet) at Columbia University (http://magnet.c2b2.columbia.edu), Columbia researchers have contributed a number of valuable new structural and systems biology algorithms to geWorkbench. geWorkbench, written in Java for use on the desktop, is open-source and cross-platform. It has strong capabilities in both analysis and visualization of data, with numerous components for microarray gene expression, sequence, and protein structure analysis. It provides direct access to numerous external data sources. Those supported by caBIG® include caArray, the Cancer Gene Index (gene-disease-compound associations), the Cancer Gene Atlas, BioCarta and the Pathway Interaction Database. MAGNet systems biology components in geWorkbench center on the creation and application of molecular regulatory interactomes. ARACNe can be used to reverse-engineer molecular regulatory networks, and MINDy is used to identify modulators of transcription regulation. The results of several such efforts are made available through the Cellular Network Knowledge Base (CNKB), which allows sophisticated searching of both experimental interactomes as well as public interaction databases. With Master Regulator Analysis, differential gene expression is analyzed in the context of an interactome to discern master regulator genes. A number of new MAGNet

components are oriented towards protein structure. MarkUs is an analysis pipeline for structure-based annotation, while Pudge integrates tools used at different stages of the structure prediction process. SkyLine and SkyBase provide homology modeling and a homology model database. To enhance the use of these tools, another MAGNet contribution is Genspace, a social networking tool which infers patterns of use (putative workflows) of geWorkbench components, and allows query and display of these real-world workflows.

AUTHORS AND AFFILIATIONS: Kenneth Smith, Ph.D.[1]; Zhou Ji, Ph.D.[1]; Thomas Garben, M.S.E.[1]; Michael Honig[1]; Oleg Shteynbuk, M.S.[1]; Udo Tibben, Ph.D.[1]; Meng Wang, M.S.E.[1]; Min You, M.S.[1]; Aris Floratos, Ph.D.[1] [1]Center for Computational Biology and Bioinformatics, Columbia University.

52. caIntegrator2 - A Translational Research Tool for 21st Century Biomedicine

caIntegrator2 is a web-based software package that allows researchers to set up custom, caBIG®-compatible web portals to conduct integrative research, without requiring programming experience. These portals bring together heterogeneous clinical, microarray, and medical imaging data to enrich multidisciplinary research. Using caIntegrator2, researchers can execute, save, and share queries to identify and collect many types of data, combining clinical information with genetic and genomic data to enable multidimensional analysis. Users can also run advanced queries, perform correlative outcome analysis using Kaplan-Meier plots, gene-expression plots, and other tools. caintegrator2 uses caGrid analytical services such as GenePattern and BioConductor to perform analysis on the integrated study data, including copy number analysis. caIntegrator2 leverages the Cancer Data Standards Registry and Repository (caDSR) to map experimental data to well defined datatypes and utilizes caGrid and Java client APIs to access data from caBIG® applications such as caArray, the National Biomedical Imaging Archive (NBIA). caIntegrator is also integrated with caBIO (Cancer Bioinformatics Infrastructure Objects) to perform queries on genes and pathways. Data access in caIntegrator2 is controlled by role-based authorization, permitting portal data to be either public or private. caIntegrator2 is an open-source, web-based J2EE Java application employing Spring, Struts 2, and Hibernate technologies.

AUTHORS AND AFFILIATIONS: Anand Basu[1]; Juli Klemm, Ph.D.[1]; Mervi Heiskanen, Ph.D.[1]; Shine Jacob[2]; Karen Ketchum, Ph.D.[2]; JP Marple[3]; TJ Andrews[5]; Ngoc Nguyen[6]; Huaitian Liu, Ph.D.[4]; Yuri Kotliarov, Ph.D.[1]; Michael Flanigan[2]; Quy Phung,Ph.D.[2]; Henry Schaefer,Ph.D.[2]; Jill Hadfield, Ph.D.[1]; Chris Piepenbring[3]. [1]NCI CBIIT; [2]ESAC; [3]5AM Solutions; [4]SAIC;[5]ScenPro; [6]ClarisLLC.

53. caNanoLab: Developing a Collaborative Environment Supporting the Application of Nanotechnology in Biomedicine

The application of nanotechnology in cancer promises advancements in early detection, targeted therapeutics, and cancer prevention and control. The use of nanotechnology in biomedicine involves the engineering of nanomaterials to act as therapeutic carriers, targeting agents, and diagnostic imaging devices. To assist in expediting and validating the use of nanomaterials in biomedicine, the NCI Center for Biomedical Informatics and Information Technology (CBIIT), in collaboration with the NCI Nanotechnology Characterization Laboratory (NCL) and other Cancer Centers of Nanotechnology Excellence (CCNEs), has developed a data sharing portal called caNanoLab. caNanoLab provides access to experimental and literature curated data from the NCL, Washington University, and other CCNEs. caNanoLab facilitates data sharing via the use of caBIG® technologies (caGrid) enabling semantic interoperability and data exchange between other caBIG® tools. caNanoLab is based on a

nanotechnology object model (nano-OM), which acts as a standard representation of nanomaterials and their physical (e.g. size, molecular weight) and in vitro (e.g. cytotoxicity, immunotoxicity) characterizations. The nano-OM leverages and extends concepts from the NCI's Enterprise Vocabulary Services (EVS) and the Nanomaterial Ontology (NPO) designed by Washington University. The nano-OM provides a model for representing the composition of diverse nanomaterial types (e.g. dendrimer, fullerene, quantum dot, carbon nanotube) and associated functionalizing entities (small molecules, antibodies). These functionalizing entities allow particles to achieve the desirable therapeutic or diagnostic functions and enable personalized medicine via the administration of targeted therapies. The project is expanding to include support for in vivo characterizations of nanomaterials and their functionalizing entities, which are analogous to those required for small molecules and other medical devices. These characterizations involve rigorous testing to determine toxicity and pharmacokinetics properties. The caNanoLab project is collaborating with members of the biomedical nanotechnology community through the caBIG® Nano WG in the development of nano-TAB, a standard supporting data import/export between disparate nanotechnology systems.

AUTHORS AND AFFILIATIONS: Juli Klemm, Ph.D.[1]; Anand Basu[1]; Piotr Grodzinski, Ph.D.[2]; Krzysztof Ptak, Ph.D.[2]; Anil Patri, Ph.D.[3]; Marty Fritts, Ph.D.[3]; Sharon Gaheen, M.B.A.[4]; Sue Pan[4]; Thai Le[4]; Elizabeth Hahn-Dantona, Ph.D.[5] [1]NCI CBIIT; [2]NCI OTIR; [3]Nanotechnology Characterization Laboratory; [4]SAIC; [5]Lockheed Martin.

54. caBIO in GeneAnswers: Integrated Interpretation of Genes

It is not enough for most researchers to identify a group of interesting genes from genomic scale experiments. People expect more advanced analysis of function and pathway for the given genes. Our team has developed a new Bioconductor package, GeneAnswers, written by R language, to automatically present potential correlation between genes of interest and specified categories based on statistical test. Besides Gene Ontology, Pathway Ontology (KEGG from Japan and REACTOME from EBI), and Disease Ontology, which the latter is derived from NCI UMLS and being developed by our group and collaborators, the caBIO pathways, integrating NCI-Nature curated Biocarta and Reactome, is also supported by the current version GeneAnswers. The up-to-date caBIO pathway queries allow to perform the pathway enrichment test in GeneAnswers and identify the potential relationship between given genes and the novel findings in biology and medicine. The package GeneAnswers can not only visualize the network between potential concepts (functions, pathways, diseases, etc.) and interested genes, but also uniquely combine optional data matrix (such as gene expression profile) and relative concepts together. With the support of Entrez eUtils, users can find the links connecting the given genes to user-defined keywords so that possible biomarkers could be identified for validation. Moreover, homologous gene mapping function of GeneAnswers make it possible for piloting experiments based on animal models to apply the achievements on human being. Furthermore, computer generated multi-concepts-genes table is introduced to integrate concepts analyses for correlated groups of genes (such as time series biological process). The package GeneAnswers, combining cutting edge caBIO pathway system and NCI based Disease Ontology with Gene Ontology, KEGG Database and REACTOME pathways, is truly helpful for physicians and researchers to gain an insight into clinical diagnosis and remedy at molecular and genetic levels.

AUTHORS AND AFFILIATIONS: Gilbert Feng, Ph.D.[1]; Pan Du, Ph.D.[1]; Kibbe A Warren, Ph.D.[1]; Simon Lin, M.D.[1] [1]Northwestern University Clinical and Translational Sciences Institute.

55. Cancer Bench-to-Bedside (caB2B): Current State and Future Directions

Cancer Bench-to-Beside (caB2B) is an open-source and secure suite of applications that permits translational research scientists to search and combine data from virtually any data service on caGrid based on semantics. The current version of caB2B suite is composed of three core components: the Web Application, the Client Application and the Administrative Module. The caB2B Web Application provides query templates that allow easy search and retrieval of microarray data (from caArray), imaging data (from the National Biomedical Imaging Archive (NBIA)), specimen data (from caTissue Suite), and nanoparticle data (from caNanoLab) across the grid. The caB2B Client Application is a thick Java application that enables advanced end users to create and execute queries across caGrid data services. The Administrative Module provides a graphical user interface for customizing a local instance of caB2B. As NCI Center for Biomedical Informatics and Information Technology (NCI CBIIT) is implementing a Semantic Service Oriented Architecture (sSOA) using the HL7 Services-Aware Interoperability Framework (SAIF), our team at Georgetown University is undertaking several activities to develop the next generation caB2B. First, to improve the functionality of the new caB2B, we are continuously reaching out to several stakeholders within and outside of the caBIG® community in order to identify unique translational research use cases. Second, to better align the new caB2B with new sSOA infrastructure, we are following the new semantic infrastructure and technology development initiatives and will provide requirements to the relevant projects being developed by the caBIG® community. Finally, to mitigate future integration challenges between new caB2B and NCI Enterprise Services (NES), we are closely watching the service specification, development, and refactoring efforts. Our poster will summarize current state and future plans for caB2B.

AUTHORS AND AFFILIATIONS: Baris E. Suzek, M.S.[1]; James B. Humphries, B.Sc.[1]; Karen Ketchum, Ph.D.[2]; Andrew Shinohara, M.S.[2]; David Tanenbaum, Ph.D.[2]; Chris Pipenbring, B.Sc.[3]; Stephen A. Goldstein, B.A.[4]; Subha Madhavan, Ph.D.[1]
[1]Georgetown University; [2]E-SAC, Inc.; [3]Capability Plus Solutions, LLC; [4]Sapient Corporation.

56. DDN: A caBIG® Analytical Tool for Comparative Network Analysis

DDN (Differential Dependency Network) is a caBIG® (cancer Biomedical Informatics Grid®) analytical tool with user-friendly interface for identifying gene network topological changes. DDN detects statistically significant topological changes in the transcriptional networks between two biological conditions and visualizes the network topological changes for further analysis. Gene regulatory networks are context-specific and dynamic, DDN uses local dependency model to represent the specific network structures under certain biological conditions by a set of conditional probabilities. The learning of the local dependency models is efficiently solved by Lasso technique. Local dependency models with unbalanced probabilities under two biological conditions preserve all possible network topological changes. To assess the statistical significance of the changed local structures, DDN carries out permutation test to assign every local structure a p-value. Local structures with p-values smaller than the threshold compose the differential dependency network. The visualized results will highlight the genes and connections that involved in the network topological changes. In the applications, DDN can accurately detect all the genes with network topological changes on simulation datasets. DDN has been applied to the estrogen-dependent T-47D estrogen receptor-positive (ER+) breast cancer cell line datasets and produces biologically meaningful results with straightforward illustrations. As a general measurement of network

topological change, DDN is extended to the evaluation of network distance between conditions or different phenotypes. In the application of identifying origin of a subtype of ovarian cancer, high grade serous carcinoma (HG) by gene regulatory network topology comparison, network distance results generated by DDN have shown that fallopian tube is the most likely origin of HG among three candidate origins due to the largest network commonality between it and tumor. The finding is consistent with newly acquired evidence.

AUTHORS AND AFFILIATIONS: Bai Zhang[1]; Lu Jin[1]; Ye Tian[1]; Jianhua Xuan, Ph.D.[1]; Yue Wang, Ph.D.[1]; Huai Li, Ph.D.[2]; Ie-Ming Shih, Ph.D.[3]; Subha Madhavan, Ph.D.[4]; Robert Clarke, Ph.D., D.Sc.[4]; Leena Hilakivi-Clarke, Ph.D.[4]; Eric P. Hoffman, Ph.D.[5] [1]Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University; [2]Bioinformatics Unit, RRB, National Institute on Aging; [3]Departments of Gynecology/Obstetrics, Pathology and Oncology, Johns Hopkins University School of Medicine; [4]Lombardi Comprehensive Cancer Center and Department of Oncology, Georgetown University; [5]Research Center for Genetic Medicine, Children's National Medical Center.

57. Management of Clinical Study Biospecimens by Integration of caTissue and ClinPortal, A Locally Developed Clinical Data Management System

Washington University School of Medicine has implemented caTissue Suite, a Cancer Biomedical Informatics Grid (caBIG®) application designed to manage the complexities of basic biospecimen annotation and tracking. However, caTissue does not provide the full functionality of a clinical data management system (CDMS); therefore, we have developed an independent tool, ClinPortal, which accommodates the complex clinical study data that are often associated with biospecimen collections. ClinPortal is an electronic case report form builder, management, and data collection tool that generalizes principles of caBIG® to facilitate collection of and access to data. Features of ClinPortal include a core UML model where each participant may be associated with one or more clinical studies, which then may be associated with one or more clinical study events that may be linked to one or more data entry forms or 'annotations.' Thus, data may be captured on each participant in the context of a specific clinical study's event point. Like caTissue, ClinPortal uses a web browser to store and retrieve data from a relational database. Its open application programmer's interface permits customized access to all application features and data integration or migration from other clinical or research data systems. Discrete pathology and clinical data entry are supported through customized data form creation. Both applications have been customized to offer deep user-interface integration for both clinical data and associated biospecimen data entry and retrieval, by linking collection protocol (caTissue) and clinical study (ClinPortal) classes with corresponding events in both applications. Together, these tools support role-based access to administrative functions, biospecimen accessioning, clinical data entry, and investigator queries. The integrated set of applications is sufficiently scalable and configurable for broad deployment across biorepositories and/or clinical studies of varying size and function, and is used in production, institution-wide, by multiple and diverse disease-focused research groups at our institution.

AUTHORS AND AFFILIATIONS: Mark Watson, M.D., Ph.D.[1]; Poornima Govindrao, M.S.[1]; David Mulvihill, M.I.M[1]; Amy Brink, B.S.[1]; Vicky Holtschlag, B.S.[1]; Dianne Oliver, M.I.M[1]; Mary Uhlmansiek M.S.[1]; Leslie D. McIntosh, Ph.D., M.P.H.[1]; Rakesh Nagarajan, M.D., Ph.D.[1] [1]Washington University School of Medicine.

58. High-throughput Biorepository Tools

Our goal in the development of biorepository tools is to support high-throughput biospecimen banking. The optimization of workflows will greatly reduce costs and time-consuming tasks involved

with the collection, processing, and data management of the large number of samples needed for discovery and validation studies. As part of this effort, we have maximized the functionality of caTissue Suite by utilizing one caTissue instance at Indiana University to support several departments and research organizations such as Pediatrics, Obstetrics, Ophthalmology, Pathology, the Indiana Clinical and Translational Sciences Institute, and the Fairbanks Institute. To date, we have 96 active protocols and data on 120,000 frozen specimens in our production instance. Due to the high specimen volume, our focus has been to automate time-consuming tasks. For this purpose we have designed and implemented several tools: 1) caTrack, which is an intelligent barcode-based automatic data capture system; 2) a sample collection scheduling feature for organized and accurate tracking of biospecimens in complex clinical studies; 3) an anticipated event design feature which creates all specimens and aliquots due to be collected over the course of a study; 4) automated storage assignment; 5) CaCore, an innovative XML-based data import and export software program; and 6) the use of scalable, globally-unique specimen identifiers. In addition to biospecimen collection clinical research requires standardized high quality clinical data collection. To enable streamlined, and cost-effective data collections, we have integrated case report form scanning and optical recognition technology with catissue Suite. We designed and implemented a method for importing data generated from scanned forms into the caTissue system and its dynamic extension tables. To efficiently accumulate large specimen collections with high quality clinical data at affordable costs, biorepository tools need to take into account all of the workflows involved and utilize time and cost reducing measures whenever possible.

AUTHORS AND AFFILIATIONS: Susanne Ragg, M.D.,Ph.D.[1,2]; Amber McMahon[3]; Sharmila Jothirajah[2]; Mahesh Nalkande[4]; Preeti Lodha[4]; Sarita Tony[3]; Fernanda Rankin[3]; Jeanne Monday[3]; Taru Jain[4]; Rachita Yadav[4]; Madhurima Bhattacharjee[4]; Abhijeet Kashnia[4]; Vipin Bansa[4]; Sameer Pendharkar[4]; Sandeep Sarkunde[4]; Kalyani Bhadekar[4]; Gunther Schadow, M.D., Ph.D.[1,2] [1]Melvin and Bren Simon Cancer Center, Indiana University; [2]Regenstrief Institute; [3]Indiana University; [4]Persistent Systems.

59. Establishing the Connectivity Between Biospecimens and Clinical Data for Enabling Translational Research and Biomedical Discoveries

Access to appropriately collected and well-annotated biospecimens is a critical requirement for researchers to enable the linkage of complementary clinical and biological data that can be used for predicting disease phenotype from molecular and genetic profiling. In this poster we will discuss current biomedical informatics efforts at the Dana-Farber Cancer Institute entailing an institute-wide adoption of caTissue including adaptations to meet functional requirements and integration with i2b2 (an informatics platform that enables utilization of clinical data for discovery research) for querying and requesting specimens.

AUTHORS AND AFFILIATIONS: Erica Woulf, BSCS, M.B.A.[1]; John Orechia, M.B.A.[1]; Jomol P. Mathew, Ph.D.[1] [1]Clinical and Translational Informatics, Dana-Farber Cancer Institute.

60. Tracking Biolgical and Environmental Samples for Large Population Based Epidemology Studies

Accurately and consistently annotating collected biological and environmental samples throughout their lifecycle from collection through exhaustion is integral to epidemiologic and translational research. The consideration of sample provenance, such as method of collection or time to processing, in combination with analytic outcomes can provide useful information on sample integrity and validity of analytic results. We describe the use of a sample

tracking system that provides comprehensive tracking of biological and environmental samples and associated data from collection through exhaustion or destruction and facilitates samples annotation. The system supports highly complex cohort studies that collect multiple and diverse biological and environmental samples from hundreds of thousands of participants and structures (home, school, workplace) at multiple time points, in multiple geographic locations, with often different collection, processing, shipment, and storage protocols for each participant and/or structure at each time point and/or location. The system also tracks analysis and dissemination of samples linking analysis results to participants, to the study that collected the sample, and to the research study that used the sample. Mapping the described tracking system to the caBIG® standards will ensure that information about collection, processing, storage, dissemination, and QA/QC is compliant with existing standards reflected in caBIG®. We provide examples from several population-based studies that illustrate the features of the biological and environmental sample tracking system and discuss our approach to attaining caBIG® bronze-level compatibility.

AUTHORS AND AFFILIATIONS: Barbara O'Brien, M.P.H.[1]; Arti Varanasi, Ph.D.[1]; Danielle Carrick, Ph.D.[1]; Charles Lawrence, Ph.D.[1]; Kellar Wilson[2]; Jerome Wernimont[2]. [1]Health Studies Area, Westat; [2] Computer Systems and Applications, Westat.

61. An Interoperable Framework to Link Biospecimen With Gene Array Data: A caArray and caTissue Integration Scenario

In the post-genomics era, biospecimens are assuming a prominent role. They provide a bridge between emerging molecular information and clinical information, by enabling researchers to study the molecular characteristics of the actual disease, and then correlating those patterns with what is known about the clinical progression of the disease. We propose a mechanism that links biospecimen and participant information in caTissue with gene array data housed in caArray to define an interoperable workflow. System Design and Implementation: In the Myeloma Institute at the Winthrop P. Rockefeller Cancer Institute at the University of Arkansas for Medical Sciences, biospecimens are collected according to IRB approved active protocols and caBIG® caTissue is being used to enter participant demographics and basic annotations about the collected specimens, with some specimens queried by microarray analyses. Once the array lab receives the specimen(s), they use an internally developed webpage to enter the specimen information. The page queries caTissue and proposes a file name that includes participant ID, specimen ID and the chip type. The proposed file name is then used to structure the file name for the Affymetrix platform. Once the test is completed, the platform generates .CHP, .DAT, .CEL, and .JPG files with the proposed file name. The generated files are housed in caArray and the researchers are given access to caArray to access array files to run their analysis. Conclusion: With the help of the proposed interoperability workflow, analyses and correlation of expression data with clinical and specimen data, can be performed with minimal effort. Since the manual entry of the specimen and participant information in array lab is eliminated, the potential data entry and mislabeling errors are drastically reduced. In addition, the lab staff time required for manual specimen linkage with the array data is reduced, allowing for more specimen analysis time.

AUTHORS AND AFFILIATIONS: Umit Topaloglu, Ph.D.[1]; Rinku Saha[1]; Jiang Bian[1]; Clyde Bailey[1]; Cheryl Lane[1]; William R. Hogan, M.D.[1]; Laura F. Hutchins, M.D.[1]; Thomas Kieber-Emmons, Ph.D.[1]; John Shaughnessy Jr., Ph.D.[1] [1]Winthrop P. Rockefeller Cancer Institute at the University of Arkansas for Medical Sciences.

62. UCSF-RTOG Cooperative Group Biospecimen Banking

RTOG (Radiation Therapy Oncology Group) is a nationally recognized clinical cooperative group and is involved in studying ways to increase the survival as well as improving the quality of cancer patients' life through clinical trials. University of California San Francisco (UCSF) is the current RTOG Biospecimen Resource location. UCSF is accredited by the College of American Pathologists (CAP) and The Joint Commission on Accreditation of Healthcare Organizations (JCAHO). This Biospecimen Resource is securely located in a building at the UCSF Mt Zion Campus with access limited to Biospecimen Resource personnel. So far in our project, we have migrated over 200K samples from a legacy system to caTissue Suite, caBIG®'s robust biorepository tool for biospecimen inventory management, tracking, and annotation. The implementation of caTissue Suite within the RTOG Cooperative Group Biospecimen Resource will help this national group manage and retrieve data concerning the collection, storage, quality assurance, and distribution of biospecimen. Current plans include integration for automated reporting to the NCI Group Banking Committee version 2 Reporting tool; the development of an interface between caTissue Suite and Aperio Digital Pathology environment; and its interoperability with caBIG®'s B2B..

AUTHORS AND AFFILIATIONS: Sudeep Basnet, M.S.[1]; Zhen Wu, M.S.[1]; Alyssa Lalanne, B.S.[1]; Mark Bridge, M.S.[1]; Sandy DeVries, M.A.[2]; Meredith West, B.S.[2]; Fred Waldman, M.D., Ph.D.[2]; Sorena Nadaf, M.S., M.MI.[1] [1]Helen Diller Family Comprehensive Cancer Center, Translational Informatics Program, University of California San Francisco; [2] Helen Diller Family Comprehensive Cancer Center, University of California San Francisco.

63. University of California San Francisco Helen Diller Family Comprehensive Cancer Center (UCSF-HDFCCC) Tissue Core Biospecimen Banking Effort

The primary function of the Tissue Core is to provide a useful tissue resource (tissue bank) to the Cancer Center and other investigators in order to support their research programs and to aid in the design, development, and successful completion of projects that involve the use of human tissue. Services provided by the Core include tissue procurement, storage, processing, and distribution. Core personnel are also involved in study design and the selection of appropriate cases for projects. Recently caTissue Suite v1.1.1 has been implemented within the core as a Biorepository tool to support the biospecimen management. The biorepository is divided into organ type specific banks. Currently the core is involved in migrating legacy data from existing core databases into caTissue Suite. The production instance of caTissue Suite application is housed in our secure Data Center with industry standard backup and security characteristics. It has been planned to connect the caTissue data with the existing Cancer Registry database at the Cancer Center. The general structure of the biorepository and the integration with already existing UCSF databases will be depicted on the poster.

AUTHORS AND AFFILIATIONS: Sudeep Basnet, M.S.[1]; Zhen Wu, M.S.[1]; Alyssa Lalanne, B.S.[1]; Mark Bridge, M.S.[1]; Hubert Stoppler, Ph.D.[2]; Sorena Nadaf, M.S., M.MI.[1] [1]Helen Diller Family Comprehensive Cancer Center, Translational Informatics Program, University of California San Francisco; [2] Helen Diller Family Comprehensive Cancer Center, University of California San Francisco.

64. The GBC Reporting Tool: A Web-Based Catalog of NCI Cooperative Group Trial Biospecimens

The NCI Cooperative Group mechanism performs multi-institutional, Phase II / III therapeutic trials, which are often associated with biospecimen collection. Although biospecimens may be intended for primary correlative science studies associated with the specific trial, a majority of the collected tissues and biofluids are also available for secondary retrospective studies and may be requested by the scientific community at large. A past limitation to more generalized use of this unique and highly valuable biospecimen resource has been the lack of a comprehensive inventory across all nine NCI-funded cooperative groups and their corresponding biorepositories. The Group Bank Reporting Tool (GBC- RT) is a web-accessible data warehouse that integrates the inventories of each cooperative group bank into a single data source. The GBC-RT uses the Common Biorepository Model v0.9 to ensure future caBIG® compatibility and connectivity with NCI Enterprise Data Services and related efforts such as the Specimen Resource Locator II. Each group bank uses a web-based tool to map vocabularies and load data from their diverse, local data systems into a common data warehouse. A web based query interface allows authorized users to search for the appropriate cooperative group bank resource(s) based on specimen and/or trial characteristics. When fully operational, the GBC-RT will be connected with the NCI Cooperative Group Bank public website and will greatly facilitate and accelerate the search and utilization of this highly valuable biospecimen resource for novel cancer research studies.

AUTHORS AND AFFILIATIONS: Dave Billiter, M.B.A,. P.M.P.[1]; Mark A. Watson, M.D., Ph.D.[2] [1]The Research Institute at Nationwide Children's Hospital; [2]Washington University School of Medicine and The NCI Group Banking Committee Informatics Subcommittee.

65. Interoperable Biorepositories, Heterogeneous Systems, and Multi-Institutional Collaboration

CaTissue Suite is utilized at our institutions to leverage the data sharing and discovery capabilities of the federated architecture of the caGrid. Internally at some of our sites a commercial system for managing and tracking its biospecimens is deployed: Daedalus Software, Inc.'s BTM-Research. To accomplish our collaborative goals, we are each faced with the challenge of moving data from an internal system to a GRID-accessible instance of caTissue Suite. Together we have developed a general strategy, and working with BTM developers we have defined an XML schema and a pathway for moving data from BTM to caTissue Suite. Our poster will highlight one implementation of our strategy: BTM is being used by the UCLA Translational Pathology Core Lab (TPCL) to catalog all biomaterials that it collects and processes on behalf of researchers at UCLA. Typically investigators develop and maintain supporting data systems and information on the patients whose biomaterials they acquire through the TPCL. These databases are external to BTM. Our poster presents an overview of the approach we have taken to merge data from several sources and import them into caTissue Suite. The movement of data begins with a built-in shopping cart metaphor in BTM. A BTM user can search for specimens and add them to a cart. The user can then click an Export to XML button that generates a specially formatted XML file that includes data about the selected biospecimens, including the external IDs that link them back to BTM. The format of this exported XML file closely matches the caTissue object model. Our group has developed a web application to insert this file into a specified caTissue Suite instance using the caTissue API.

AUTHORS AND AFFILIATIONS: Andrew Helsley[1]; Stefano Santoro[2]; Azita Sharif[2]; Paul Fearn[3]; Nick Anderson, Ph.D.[3]; David Steffen, Ph.D[4]; Robert A. Dennis, Ph.D.[1] [1]Computing Technologies Research

Lab, David Geffen School of Medicine at UCLA; [2]Daedalus Software, Inc.; [3]Biomedical and Health Informatics, University of Washington; [4]Bioinformatics at Baylor College of Medicine.

66. Continued Development of caTIES, A Collaborative Tissue Banking and Text Mining Tool

We report on the continued development of the Cancer Tissue Information Extraction System (caTIES): an application that supports collaborative tissue banking and text mining by leveraging existing natural language processing methods and algorithms, grid communication and security frameworks, and query visualization methods. The system fills an important need for text-derived clinical data in translational research such as tissue-banking and clinical trials. The design of caTIES addresses three critical issues for informatics support of translational research: (1) federation of research data sources derived from clinical systems; (2) expressive graphical interfaces for concept-based text mining; and (3) regulatory and security model for supporting multi-center collaborative research. Implementation of the system at several Cancer Centers across the country is creating a potential network of caTIES repositories that could provide millions of de-identified clinical reports to users. The system provides an end-to-end application of medical natural language processing to support multi-institutional translational research programs. caTIES was originally funded under the caBIG® program and is now supported by a five-year NCI R01 continued development grant whose goals are: (a) improving the portability of the system and extending the types of documents that can be processed; (b) evaluating the system's NLP performance and usability; (c) building a user community to support this open-source application; and (d) piloting interoperability of caTIES with other enterprise and research systems.

AUTHORS AND AFFILIATIONS: Girish Chavan, M.S.[1]; Kevin Mitchell, M.S.[1]; Yining Zhao, M.S.[1]; Rebecca Crowley, M.D., M.S.[1] [1]Dept.of Biomedical Informatics, University of Pittsburgh School of Medicine.

67. Connecting TissueMetrix to the caBIG® Using the CBM Database and Web Service

In mid-2009, the Hollings Cancer Institute (Charleston, SC), with the assistance of Artificial Intelligence In Medicine Inc. (Toronto, Canada), undertook the challenge of connecting its existing biospecimen repository management system with the Cancer Bioinformatics Grid (caBIG®) via the Common Biorepository Model (CBM) and software components available from the National Cancer Institute's (NCI). This poster describes the motivation and objectives of the project, connection architecture, the steps involved, technical challenges and solutions, and the effort involved in mapping data elements and values from the local controlled vocabulary to the CBM terminology. The poster also discusses challenges related to the ongoing maintenance and upkeep of the grid presence. This poster will be of interest to those who wish to integrate existing research biospecimen databases with the cancer bioinformatics grid. Exposing biospecimen information on the grid can lead to renewed collaboration and accelerate research activities but requires, among other things, attention to operational policies, procedures, and data sharing agreements between institutions.

AUTHORS AND AFFILIATIONS: William Morgenweck[1]; Kevin Hooper[2]; Lucy Gansauer,[3] R.N., M.S.N., CHSP, CPSO, OCN.[3] [1]Hollings Cancer Center; [2]Artificial Intelligence in Medicine; [3]Spartanburg Regional Healthcare.

68. Developing a Custom User Interface for Real-time Data Entry into caTissue

Fox Chase Cancer Center's (FCCC) Biosample Repository collects blood and tissue from thousands of participants each

year. Biospecimen processing staff maintain an up-to-date inventory of available samples in this critical, heavily used facility. The primary objective for a new Biosample Repository Inventory System was to create an easy to use application. The previous version of the inventory system, developed over 10 years ago, had an adequate data model and comprehensive feature set. However, Biosample Repository staff found some functionality cumbersome and instead relied on external tools, such as Microsoft Excel and Access, to meet certain needs (e.g., recording processing and shipping information). Multiple data sources with potentially conflicting information made integration of inventory data with patients' medical records difficult and time consuming. Together with end users we evaluated both commercial and open source tissue banking software. None of the evaluated products could support our workflow without significant modification. After comparing costs of deployment and customization of different solutions, we elected to develop a new interface for caTissue. We implemented our own object-relational mapping on top of the caTissue data model and added a service layer to match our workflow. We used Google Web Toolkit (GWT) to create an Excel-like user interface and deployed it as part of our Liferay-based research portal, integrating it with Enterprise Master Patient Index (EMPI) and other FCCC study management portlets. Extending caTissue allowed us to create an application customized to FCCC biospecimen processing workflow, yet compatible with caBIG® standards and ready for data sharing via caGrid. The user-friendly, spreadsheet-like interface for real-time data entry improved user compliance and helped establish a single reliable source of inventory information.

AUTHORS AND AFFILIATIONS: Olga Tchuvatkina, M.S.[1]; Yan Huang, M.A.[1]; Michael A. Collins, M.S.[1]; Eric A. Ross, Ph.D.[1]; J. Robert Beck, M.D.[1] [1]Fox Chase Cancer Center.

69. Adoption and Adaptation of caTissueSuite at the University of Iowa

In 2008, in the course of a comprehensive update of its tissue procurement core, the University of Iowa's Holden Cancer Center (HCCC) adopted caTissueSuite for bio-repository management. During the initial process of testing it became evident that modifications would have to be made to the base software configuration to support production use. Developers at the University of Iowa both modified the caTissueSuite code base and wrote a suite of complementary Perl/cgi modules to accomplish this goal. The code base was modified to provide additional validation for participant enrollment and to remove some of the features from the original GUI design that were not needed by the tissue core. The Perl code enables the expedited entry of specimen-collection-group and specimen information, uploading of consent files, and the conversion of plain text files into HL7 v.2.3.1 for the caTIES-like functionality in caTissueSuite. Currently additional Perl/cgi modules are being written to expedite end-of-the-month reports. Other activity includes the assessment of the de-identification software HIDE, developed at Emory, for use in de-identifying surgical pathology reports.

AUTHORS AND AFFILIATIONS: Bartley Brown, Ph.D.[1]; George Weiner, M.D.[2]; Terry Braun, Ph.D.[3]; Thomas Casavant, Ph.D.[1] [1]The Center for BioInformatics and Computational Biology, The University of Iowa; [2]The Holden Comprehensive Cancer Center, The University of Iowa; [3]The Coordinated Laboratory for Computational Genomics, The University of Iowa.

70. Sharing Biospecimen Inventory From Locally-Developed Databases Using the CBM

The University of Colorado Cancer (UCCC) encompasses multiple biorepositories facilitating the processing and storage of hundreds of thousands of cancer-related biospecimen from many disease sites. These specimens are tracked using several

well-accepted, locally-developed, heterogeneous, biorepository databases. To enable sharing of specimen inventory across the research enterprise "both within and outside the University" UCCC is implementing the caBIG® Common Biorepository Model (CBM). The CBM provides a streamlined, standardized model for sharing biospecimen inventories. While it was primarily designed for biorepository database vendors, UCCC believes it is equally well suited for use with locally-developed databases. UCCC surveyed the various campus biorepository databases to determine their readiness to share data via the CBM. Based on these findings, target biorepositories were selected for a pilot roll-out of the CBM. Requirements and technical designs were developed and a proof-of-concept implemented. This poster will discuss the merits and challenges of implementing the CBM in the context of a locally-developed biorepository database, methods used by UCCC in determining a strategy for a pilot roll-out, technical issues encountered, solutions implemented, preliminary outcomes, and plans for future development.

AUTHORS AND AFFILIATIONS: Michael J. Ames, M.B.I.[1]; Kenneth Claussen, M.B.A.[1]; Jessica Bondy, M.H.A.[1] [1]University of Colorado Cancer Center Research Informatics Core.

71. Integration of caTissue with Clinical Data in an i2b2 Research Data Mart

Biospecimens used for research must have associated quality control, pathological annotation, and clinical annotation to be of value. For clinical annotation, caTissue currently provides for participant demographics, protocol registration, and clinical diagnosis. Provision for acquiring additional clinical annotation is available through caTissue's dynamic extensions functionality. However, relying on caTissue to house complete phenotype information on specimens would not only be difficult to achieve, but would not be an appropriate information architecture strategy. Rather, clinical data warehouses are more comprehensive sources of clinical data, and combining de-identified warehouse phenotypes with caTissue biospecimen information in research data marts is a manageable and efficient means to providing investigators the ability to query the availability of specimens needed for their research. At Thomas Jefferson University comprehensive clinical data, including demographics, diagnoses, laboratory values, medications, procedures, and outcomes, from our hospital's clinical data warehouse (CDW) has been de-identified and loaded into an i2b2 research data mart (RDM). The informatics for integrating biology and the bedside (i2b2) framework is a widely used approach for integrating clinical and research data that was developed with NIH support at Partners Healthcare System. We first augmented the i2b2 ontology for the RDM with biospecimen annotation so that caTissue data could be regularly uploaded to the RDM and available for queries. Additionally, i2b2 query tool plug-in software was developed using caTissue's APIs, enabling the extraction of additional biospecimen data directly from caTissue for specified cohorts. This provides a scientist workflow in which the i2b2 query tool can be used to discover cohorts with the desired phenotype and available specimens, while the caTissue application remains the primary means of then identifying and ordering the specimens. Thus a complementary relationship exists between caTissue and i2b2. Work is also underway at several other institutions to federate i2b2 research data marts using caGrid.

AUTHORS AND AFFILIATIONS: Jack W. London, Ph.D.[1]; Karl J. Smalley[1]; Devjani Chatterjee, Ph.D.[1]; Jeremy M. Isikoff[2]; Dan Housman[2]; Matvey B. Palchuk, M.D., M.S.[2,3] [1]Kimmel Cancer Center, Thomas Jefferson University; [2]Recombinant Data Corp.; [3]Harvard Medical School.

72. Sharing Biorepository information through the Common Biorepository Model (CBM) and NCI Specimen Resource Locator

Basic life science and clinical research often use locally obtained specimens since there are no easy methods to search for specimens outside a lab or institution. The ability to aggregate similar specimens from various sites will expand the validation of research findings and thus, more quickly impact patient care. The National Cancer Institute is leveraging the caBIG® infrastructure (caGrid) and semantic/syntactic interoperability, the goals of the Office of Biorepositories and Biospecimen Research to enable faster and better access to specimens, and various stakeholders (NIH/NCI-funded biorepositories, biospecimen management system vendors, cancer centers, and other biorepositories) to develop, test, and adopt the Common Biorepository Model (CBM). Over the past year, the CBM early iterations have been provided to the community for feedback/testing and today several CBM participants have established test grid nodes with test data on the caGrid training portal, allowing caGrid querying of their biorepository data, and the use of other caBIG® tools (such as caB2B) to be used. The latest CBM1.0Beta version has a solidified set of diagnosis terms that are in the NCI Metathesaurus and was generated from the input of various Specimen Resource Locator stakeholders. In conjunction with the CBM development, the NCI Specimen Resource Locator v2.0 development is underway and will be able to provide a web interface that emphasizes ease-of-use for researchers and displays dynamically-updated biorepository information. The updated information will come from CBM caGrid services associated at individual biorepositories across the country, as well as periodic manual updates from centers.

AUTHORS AND AFFILIATIONS: A. T. Fernandez, Ph.D.[1]; I. Fore, D.Phil.[2]; A. Breychak[3]; E. Prince [3]; B. Fombonne[4]. [1]Booz Allen Hamilton; [2]NCI CBIIT; [3]Sapient Government Services; [4]Kelly Government.

73. the Goal: How Interoperability Will Support the Cancer Human Biobank (caHUB)

Background: Critical to making progress in the fight against cancer is the availability of high-quality human biospecimens. In turn, the availability of high-quality biospecimens is dependent upon standardized handling processes, removal of competitive barriers to biospecimen access, and biospecimen research science that explores how collection, processing, storage, and transport procedures impact the biospecimen's molecular characteristics and ultimate usefulness in cancer research. Because the current biobanking system has fallen short in all of these areas, there is a considerable and problematic shortage in the availability of high-quality, well documented biospecimens for cancer research. Methods: To address this shortage, the National Cancer Institute (NCI), through its Office of Biorepositories and Biospecimen Research (OBBR), is developing a national, standardized human biospecimen resource called the cancer Human Biobank (caHUB) that will serve as a continuous and reliable source of high-quality human biospecimens and associated data for the broad cancer community, including basic and clinical researchers and the biotechnology and pharmaceutical industries that rely on biospecimens for cancer diagnostics and drug development. The Cancer Human Biobank (caHUB) aims to address the problem of specimen inconsistency by setting standards for procuring, preparing, and storing specimens and for data that must be collected to describe the specimens. To support this effort, caBIG® is developing a Global Unique Specimen Identifier (GSID) service that will provide a unique identifier across the universe of collection site involved in caHUB. Results: This poster will illustrate the progress-to-date in building the

operational and informatics components of the caHUB, including the anticipated use of caBIG® tools and infrastructure. Conclusions: With the help of the interoperability provided by the caGRID services and caBIG® tools and methodology, caHUB hopes to increase the integrity of medical research, but also pave the way for personalized medicine, the healthcare paradigm for the 21st century.

AUTHORS AND AFFILIATIONS: I Fore, D.Phil.[1]; K. Erickson[2]. [1]NCI CBIIT; [2]Sapient Government Services.

74. Harmonizing the National Institute of Neurological Disorders and Stroke (NINDS) Core Common Data Elements (CDEs) with the National Cancer Institute (NCI) Cancer Biomedical Informatics Grid® (caBIG®) Cancer Data Standards Registry and Repository (caDSR)

In an effort to reduce study start-up time and accelerate data sharing in neurological clinical research, the NINDS started the CDE Project four years ago. This effort has resulted in the development of General core CDEs, which are commonly collected in all clinical studies regardless of therapeutic area, and more recently, the development of disease-specific CDEs. Meanwhile, NCI invested in a system of standards and a data warehouse structure that would allow access to and contributions from researchers sponsored by the NIH Institutes and other national and international institutions. The importance of data integration is critical because data collected at different agencies, institutions, and clinics cannot be compared or analyzed in aggregate. With the goals of data harmonization and integration in mind, the NINDS CDE Team (NCT) worked with the caBIG® data standards team to perform a gap analysis of the structure and definitions of the NINDS CDEs compared to the NCI caBIG® standards. Tools from the caBIG®, specifically the CDE Browser and Curation Tool, were used to map the NINDS core CDEs to the caDSR structure. This initial effort focused on the semantic interoperability between the NINDS CDE and caBIG® vocabularies. The mapping revealed that of 151 CDEs on 21 CRFs, 123 CDEs were identified as "harmonized" or partially harmonized with caDSR standards (81%). The successful mapping of the NINDS General CDEs serves as a proof of concept so that future NINDS disease-specific CDEs and CRFs can be integrated with the caBIG® standards. The NCT team is in the process of re-evaluating the NINDS General core CDEs based on the pilot harmonization effort with caBIG® and feedback from NINDS communities. This presentation will describe the result of this re-evaluation and revision to maximize the utility and availability of the NINDS General core CDEs.

AUTHORS AND AFFILIATIONS: Yun Lu, Ph.D.[1]; Dianne Reeves, R.N., M.S.N.[2]; Stacie Grinnon, M.S.[1]; Kristy Miller, M.P.H.[1]; Patti Shugarts, B.S.[1]; Joanne Odenkirchen, M.P.H.[3] [1]KAI Research, Inc; [2]NCI CBIIT; [3]NINDS.

75. Forms Curators - What They Do and How They Do It

There are currently 12 caDSR Contexts contributing metadata from a variety of sources including caBIG® applications (UML Models), standard data dictionaries, protocols and case report forms (CRFs), questionnaires and research surveys. In order to create and maintain the metadata, there are 15 to 20 active curators (who have completed the self-paced and web-based caCore training offered by NCI CBIIT) working in at least one, but sometimes several Contexts. Using a well-established business process, curators choose the best process for metadata registration using either manual curation or UML modeling techniques, search for existing metadata content to reuse as is or modify, determine concepts to use to build new metadata, request new concepts for inclusion in the NCI Thesaurus when needed, and use tools to create data element concepts, value domains, and data elements. Curators also

create forms in Form Builder for use by Clinical Study Builders working with electronic data capture applications. Throughout this process, forms curators act as liaisons facilitating consensus through iterative reviews among stakeholders on the specification and meaning of the items contained in source documents. When consensus is achieved, the metadata is released for reuse by other caDSR community users. This community is comprised of NCI and NIH organizations, Standards organizations, research and commercial partners and universities. CDEs are important to the forms community as vocabularies (concepts) alone do not capture the contextual information about data files. Reusability of existing metadata content moves the community toward harmonization for forms and standardization of forms metadata. Forms curators are forward thinking and will support the transition to the new semantic infrastructure, minimizing disruption for the community, and positioning metadata content for the future. Forms curation will continue to be done as activities to support transition to new semantic infrastructure take place.

AUTHORS AND AFFILIATIONS: Sana Ahmed[1]; Janice Chilli[1]; Mary Cooper[1]; Tommie Curtis[1]; Kathleen Gundry[1]; Jocelyn Leatherwood[1]; Brenda Maeske[1]; Dianne Reeves[2]. [1]SAIC; [2]NCI CBIIT.

76. Finding Information on the caDSR Content Wiki

The Content Wiki focuses on the metadata content within the caDSR. This metadata information is developed by groups and organizations divided into domain-specific work areas called Contexts and led by Context Administrators. A Context Curator for each group provides oversight for creation, maintenance, and designation of common data elements (CDEs) within the Context. The Context Administrators and Curators work together as a group, known as the Content Team, to coordinate and harmonize metadata development efforts. The Content Team develops and documents business rules that pertain to both to the creation and development of metadata, as well as to the management and review of case report forms (CRFs). The team also approves best practices that outline techniques and methods for efficiently and effectively curating metadata in the caDSR, based on repeatable processes. As caDSR metadata is migrated to the new semantic infrastructure, the Content Team will work to document changes in business practices with the goal of producing standard metadata content that will easily transition to the new infrastructure. The Wiki pages include links to the current business rules and best practices, caDSR Content harmonization efforts, NCI/caBIG® data standards and governance, and the NCI Training Wiki. The Content Wiki provides a valuable tool that facilitates the dissemination of information and the review of draft information.

AUTHORS AND AFFILIATIONS: Sana Ahmed[1]; Janice Chilli[1]; Mary Cooper[1]; Tommie Curtis[1]; Kathleen Gundry[1]; Jocelyn Leatherwood[1]; Brenda Maeske[1]; Dianne Reeves[2]. [1]SAIC; [2]NCI CBIIT.

77. Connecting EVS with the Semantic Web

The caBIG® community has recognized the need for semantic annotation of data models and data since its inception. caBIG® annotation is currently done via the NCI Enterprise Vocabulary Services (EVS), which, in turn, is built on the LexEVS software suite. Over the past couple of years, the W3C Semantic Web community has become actively involved in semantic markup of data and metadata and has developed its one set of representations, tools, and de-facto standards. Here we explore the options for combining the existing LexGrid model that underlies the LexEVS services with emerging W3C standards. The LexGrid model currently provides a common set of semantics and the LexEVS loaders provide a set of transformations that map the disparate syntax and semantics of RRF, OBO, OWL,

CSV, HL7 RIM, etc to a common, shared model. The fact that this already exists puts the caBIG® community at a distinct advantage, as a single semantic map between LexGrid and W3C standards (including SKOS/OWL/RDF/Dublin Core and OMV) will allow existing terminological content and annotations to migrate forward while, at the same time, providing access to emerging W3C ontologies through the same interface. The terminological information represented in LexGrid can therefore be translated to RDF triples, and therefore allowing LexGrid to leverage standard tools and technologies for storage, querying, and reasoning capabilities. For more information about the LexRDF project, please see the following papers: 1. Tao C, et al. LexRDF Model: An RDF-based Unified Model for Heterogeneous Biomedical Ontologies. In Proceedings of the International Workshop on Semantics for Rest of Us in conjunction with the 8th International Semantic Web Conference (ISWC 2009) Chantilly, Virginia Oct. 2009; and 2. Tao C, et al. A Standard Common Terminology Model for Representing Biomedical Ontologies in RDF Triples, submitted to JBI for publication.

AUTHORS AND AFFILIATIONS: Cui Tao, Ph.D.[1] Harold R. Solbrig[1]; Jyotishman Pathak Ph.D.[1]; Christopher G. Chute MD Dr.Ph.[1] [1]Mayo Clinic.

78. Towards an Unambiguous and Formal Description of Cancer Therapy Experiments

Recording and reporting experiments - their design, context and results - in an unambiguous manner is crucial for the advancement of biomedical research, as it enables reproduction, well-grounded comparisons, reuse, and integration of data from different experiments. Thus, unnecessary repetition is avoided and reliable analysis is possible due to improved statistical power of the data. While the experiments' description should avoid different interpretations, the data exchange formats must also allow sharing, integration, and interoperability. In this work, we focus on cancer therapy experiments. Previously, some of the authors presented Guidelines on Information about Therapy Experiments (GIATE). GIATE consists of a list of Common Data Elements (CDEs), as per the ISO 11179 metadata registries' standard. Some CDEs were extracted from the caBIG® semantic infrastructure; others were created specifically. The CDEs are annotated with the NCI thesaurus ontology. We present GIATE as an ontology, or formal conceptualization, of cancer therapy experiments. As opposed to a list of information elements, the ontology supports an unambiguous and formal description. We used the Web Ontology Language (OWL), which is based on description logics and recommended by the World-Wide Web Consortium. By using OWL as data exchange format, following a semantic web/linked data approach, sharing, integration and interoperability are guaranteed. While GIATE was developed independently of the NCI thesaurus to focus on the specific sub-domain related to therapies, we also produced a matching between the two ontologies. This matching will facilitate the interoperability between GIATE-compliant knowledge bases with caBIG® data services. This poster also includes a case study demonstrating how the GIATE ontology is used to model two different experiments. It is our view that the GIATE ontology is a further step towards achieving integrative translational research and that it should undergo comprehensive inspection by the cancer research community to be used as a recording standard.

AUTHORS AND AFFILIATIONS: Alejandra Gonzàlez-Beltràn, Ph.D.[1]; May Yong, Ph.D.[2]; Richard Begent, M.D.[2] [1]Wolfson Institute of Biomedical Research and Department of Computer Science; [2]Cancer Institute, University College London.

79. Ontology-Based Queries for the caGrid® Infrastructure

Sharing, searching, and integrating data is important for the advancement of

biomedical research. To facilitate these tasks, the UK National Cancer Research Institute Informatics Initiative (NCRI II) promotes standards and tools and collaborates with the US National Cancer Institute caBIG® programme. This study presents our work at University College London on ontology-based queries over the caBIG® infrastructure, caGrid, as part of the NCRI II plan to maximise the impact of cancer research. In caGrid, data are exposed as services following a model-driven architecture: they are based on information models described by the Unified Modeling Language (UML) and annotated with concepts from the NCI thesaurus (NCIt) ontology. The annotations provide unambiguous meaning and support semantic interoperability. A metadata registry, caDSR, maintains the mappings between UML models and their semantic annotations. In this way, the NCIt ontology serves as a conceptual unified view of the data services. However, the caGrid query functionality does not currently consider this conceptual view and it is based solely on the UML models. Our approach for ontology-based queries over caGrid® is based on semantic web technologies and, in particular, the Web Ontology Language (OWL). OWL is a formal language for knowledge modeling based on description logics, whose latest version, OWL2, is a W3C recommendation since October 2009. Our approach involves: (1) module extraction from the NCIt ontology; (2) UML to OWL conversion; and (3) query rewriting from queries expressed with NCIt concepts to caGrid queries. As an extension to previous work, we developed a caGrid analytical service offering methods that, given a project in caDSR, extract modules from the NCIt and convert the UML models into OWL. Additionally, we have revised the query rewriting process in light of the OWL2 profiles and the theoretical results on the trade-off between expressiveness and computational complexity.

AUTHORS AND AFFILIATIONS: Alejandra Gonzàlez-Beltràn, Ph.D.[1]; Ben Tagger, Eng.D.[2]; Anthony Finkelstein, B.Eng., M.Sc., Ph,D., C.Eng., FIET, CTP, FBCS.[2 1] Department of Computer Science and Wolfson Institute of Biomedical Research, University College London; Department of Computer Science, University College London.

80. Curating All Data Elements In A Clinical Trial with CDEs from caDSR

The National Cancer Institute (NCI) is trying to standardize Common Data Elements (CDEs) used in cancer research data capture and reporting. The common structure will enable collaborators to develop a coherent set of CDEs suitable for sharing. To overcome the present challenges and obstacles for such standardization, NCI has developed the ISO/IEC 11179 based Cancer Data Standards Repository (caDSR). The Winthrop P. Rockefeller Cancer Institute at the University of Arkansas for Medical Sciences has decided to implement standard operating procedures that require use of CDEs from caDSR for all investigator-initiated trials for clinical as well as bench research data collection. For that very purpose, we have trained a staff person to be the local CDE curator to find/create CDEs at the institution. System Design and Implementation: We report a clinical trial of which all the data elements in electronic Case Report Form (eCRF) were curated and created in our Clinical Data Management System (CDMS). The trial is a device study and has 12 eCRFs. The total number of data elements in the study is 72 and 38 data elements were already created in caDSR. The remaining 33 CDEs (e.g. Anesthesia Start Time 3029420v1.0) were created by our local curator in draft/new status and waiting to be approved by the NCI and become released. Conclusion: As part of our commitment to standardization efforts, we have created the eCRFs in a CRF pool within our CDMS. In addition, we have installed caBIG® OpenMDR tool and currently it is being tested so we can automate CDE creation and reuse. We believe that such a process is going to help research data collection and collaboration

with the other researchers across the nation.

AUTHORS AND AFFILIATIONS: Umit Topaloglu, Ph.D.[1]; Tong Jett[1]; Cheryl Lane[1]; William R. Hogan, M.D.[1]; Adam Hicks[1]; Thomas Kieber-Emmons, Ph.D.[1]; Laura F. Hutchins, M.D.[1] [1]Winthrop P. Rockefeller Cancer Institute at the University of Arkansas for Medical Sciences.

81. Population-Based Data Collection Utilizing Open-Source Software and Common Data Elements

We report on an open-source infrastructure that enables the collection of high-quality, standardized survey data for research. We demonstrate its utility by re-creating the Breast Modular Mammography questionnaire at the Cancer Institute's Cancer Control Core. The most labor-intensive step was finding existing, and when necessary, creating new common data elements. Collecting epidemiological data can help to understand the effects of environmental factors on disease progression, and can be used for preventive medicine. Because survey data are a critical component of case-control, case-cohort, or similar study designs, a reliable informatics infrastructure that can house and share the data for collaboration has vital importance. System Design and Implementation: This infrastructure includes (1) LimeSurvey, an open-source application for conducting surveys; (2) common data elements (CDEs) from the National Cancer Institute; (3) import of CDEs into LimeSurvey; and (4) the free DQ Analyzer tool and other methods for ensuring the quality of the data collected. In our proposed workflow, a researcher, or a member of the research staff, works with a trained CDE curator to curate the already designed survey and then put it in LimeSurvey. He or she can then publish the survey on-line or print it for paper-based data collection. For paper-based data collection, we use an Optical Character Recognition OCR application to scan the surveys and import the data. Finally, we implemented several Information Quality (IQ) measures, such as data-quality assessment, root-cause analysis of existing issues, and development and implementation of improvement plans to minimize data entry errors and clean historical data. Conclusion: We built an open-source framework for the collection of standardized, high-quality survey data. As a result of the work, existing survey data was analyzed, corrected, standardized, and formatted to the new survey database developed for further research and data collection purposes.

AUTHORS AND AFFILIATIONS: Umit Topaloglu, Ph.D.[1]; Tong Jett[1]; Val Pushkarev[1]; Desiree Burroughs[1]; Cheryl Lane[1]; Jasmine Brown[1]; Jaqualane Scales[1]; William R. Hogan, M.D.[1]; Ronda Henry-Tillman, M.D.[1]; Thomas Kieber-Emmons, Ph.D.[1] [1]Winthrop P. Rockefeller Cancer Institute of the University of Arkansas for Medical Sciences.

82. Tools Used to Create High Quality Common Data Elements (CDEs)

The Stem Cell Act of 2005 mandated the collection of hematopoietic stem cell transplant (HSCT) data in the Stem Cell Therapeutics Outcomes Database (SCTOD). Transplant centers may fill out and fax paper forms or utilize a web-based application (FormsNet) to enter this data, resulting in data duplication and inconsistencies with local transplant center databases and information systems. AGNIS® (A Growable Network Information System) minimizes this by electronically completing HSCT forms using local data and ultimately transmitting them to the SCTOD. To facilitate the electronic data transmission, AGNIS utilizes Common Data Elements (CDEs) created using the Cancer Data Standards Registry and Repository's (caDSR) curation tools. The National Marrow Donor Program is currently curating 98 forms with approximately 14,000 data elements. Maintaining consistency within and between forms has proven to be quite challenging. To help manage this, we have set in place several tools throughout the curation process to ensure high quality

content. Several templates are used in the creation of CDEs: (1) A Local Terminology spreadsheet lists our common terms and definitions approved by our subject matter expert; (2) The Decomposition Template defines how common questions should be curated; (3) A Question Text template maintains consistency with how questions are asked; (4) A Validation Rules spreadsheet has been setup to maintain consistency with preferred DECs and VDs; and (5) the cgMDR (Cancergrid Metadata Registry) tool is used for ease of search and input capabilities.We have also created a Visual Basic application that generates reports used to review our content. Several report selections can be made, including automated error checking to review various aspects of the curation process. Once the curation and review processes are complete, we use a custom-built Java application, the AGNIS® Metadata Tool, to assist with the mapping of the CDEs to local database fields.

AUTHORS AND AFFILIATIONS: Michele Nych[1]; Alexandra Morozoff[1]; Robinette Aley[1]; Joel Schneider[1]; Kirt Schaper[1]; Barb Kramer[1]; Martin Maiers[1]; Colleen O'Neill[1]; J. Douglas Rizzo, M.D., M.S.[2] [1]National Marrow Donor Program; [2]Center For International Blood and Marrow Transplant Research.

83. caBIG® and REDCap, CTSA and UCSF HDFCCC CRI - Incrementally Advancing Translational Informatics

The Research Electronic Data Capture system (REDCap) is a robust meta-data driven rapid application development framework from Vanderbilt geared to support web-based data capture for single and multi-site research studies. As of July 2010, the REDCap Consortium had 127 active institutional partners using or developing 2,570 applications with a user base of over 8,200 research end-users. REDCap is used at UCSF HDFCCC CRI because, as in many institutions across the country, there is often too much inertia and too many special cases to go straight to a standards-based system and REDCap is powerful and flexible enough to allow for an incremental transition. REDCap projects have been launched in as little as one day and are routinely used to support small single-site studies as well as large multi-site international studies. REDCap was developed with a 'researcher first' mentality around the philosophy of giving scientists an 'easy way to do the right thing' when planning and conducting research studies. Most features and functions were developed based on direct input from research teams. For example, one recent addition to the program was a standard library of downloadable data collection instruments for immediate use by research teams creating new studies. The purpose of this project was to assess the feasibility of leveraging REDCap infrastructure to share terminology and data element work done in the caBIG® community for research teams - specifically in the cancer scientific domain. In phase I, we assessed feasibility of using the NCI Standard Demography Module Template CRF work in the REDCap consortium instrument library. We anticipate rollout of this work into the REDCap community for use in Q3, 2010. Phase III will include exploration of utilizing REDCap data in interfaces across the caBIG® Clinical Trials Suite and leveraging other enterprise services at NCI.

AUTHORS AND AFFILIATIONS: Sorena Nadaf, M.S. M.M.I.[1]; Sudeep Basnet, M.S.[1]; Mark Bridge, M.Sc., M.B.C.S.[1]; Alyssa LaLanne, B.A.[1]; Zhen Wu, M.S.[1]; Paul A. Harris, Ph.D.[2] [1]University of California, San Francisco Helen Diller Family Comprehensive Cancer Center and Cancer Research Institute; [2]Office of Research Informatics, CTSA Biomedical Informatics, Department of Biomedical Informatics, Vanderbilt University.

84. Implementation of CTCAE Version 4.0 in C3D Studies

The following topics will be discussed in this poster: (1) What's new in CTCAE v.4.0 and how is it related to MedDRA.; (2) What are the characteristics of CTCAE v.4.0; (3) Implementation of CTCAE v.4.0 for new

studies in C3D; and (4) Implementation strategy of CTCAE v.4.0 for existing studies.

AUTHORS AND AFFILIATIONS: Lily Jiang[1]. [1]NCI CBIIT.

85. caDSR Bulk Loading Process

Loading ISO 11179 entities into caDSR is a two step process. The first step requires the user to gather information required for creation of the entities from caDSR or controlled vocabularies. The second step requires a tool which interprets gathered data, transforms it into a ISO 11179 compatible format and then performs the load. The caDSR Bulk Loader creates and loads ISO 11179 entities such as Data Elements, Data Element Concepts, Value Domains, and other Administered Items including related Concepts into the caDSR in bulk, in our case from annotated Excel spreadsheet(s). The loader also provides a module for transforming exported data from an annotated Excel file into an XML document that conforms to the caDSR ISO 11179 based XML schema. A validation module performs the necessary data quality check prior to transforming and while loading the data into the caDSR. The semantic annotations on the Excel spreadsheet can be created manually or by using the cgMDR Excel add-in. The add-in provides a common interface from within Excel to search for semantic annotations from existing caDSR content, LexEVS, or concepts in the local cgMDR. The add-in also facilitates typing columns in an Excel spreadsheet with the retrieved elements for CDE-driven data collection. The cgMDR add-in has been enhanced to support a simplified click-through installation of all the components necessary to work with the caDSR Bulk Loader to enable creation of ISO 11179 metadata in the caDSR in bulk, without the need for a UML model.

AUTHORS AND AFFILIATIONS: David Hau, M.D., M.S.[1]; Denise Warzel, M.S.[1]; Denis Avdic[2]; Ashwin Mathur[2]; Steve Harris, Ph.D.[3]; Andrew Tsui[3]. [1]NCI CBIIT; [2]Ekagra Software Technologies; [3]Oxford University Computing Laboratory.

86. Safety Profiler

Safety ProfilerTM, developed by CTIS, Inc., is a wireless reporting solution built upon Windows Mobile 6.0 platform, has strong implications for improving the efficiency and completeness of toxicity data collection for clinical trials conducted at various levels including NIH/NCI, pharmaceutical industry, and locally funded studies. The features included in Safety ProfilerTM assist in identifying adverse event (AE), including toxicity assessment, AE grading, and reporting and maintains regulatory compliance the through use of M1/MedDRA, CTCAE, 21 CFR Part 11, ICH-GCP and HIPAA standards. Two primary systems features, data validation at the point of care and data standardization, are backed by AE term data libraries and CTCAE dictionaries, which are incorporated into the tool to assist users while entering data in standard language. User capabilities are further enhanced through the availability of cross-mapped full terms, short names, and medical codes that are part of CTCAE. Validation of the data as the user generates the AE information ensures that the data is relevant, caBIG® VCDE (Vocabulary and Common Data Elements) compliant, and that the report is complete leading to faster, more accurate reporting, and to more timely NCI or FDA action in response to SAE reports. The VCDE Workspace of caBIG® is responsible for standards oversight of NCI ontologies and vocabularies. The Safety ProfilerTM product also bears Microsoft's approval as a Designed for Windows Mobile application.

AUTHORS AND AFFILIATIONS: Rajiv Shahi, M.Sc., P.M.P., M.C.A.D.[1] [1]CTIS Inc.

87. Ontology Modeling of caBIG® Semantics

The semCDI methodology has been proposed to provide knowledge representation and concept-based querying within the caGrid environment using

Semantic Web standards. The ultimate objective of semCDI is to enable execution of queries against concepts in NCI Thesaurus capable of returning valid data instances from caGrid data services. The core of semCDI consists of the mechanisms designed to achieve a representation of the caBIG® semantics as ontologies formulated in the Web Ontology Language (OWL). In this poster, we present the rationale for two design choices made within semCDI. First, we discuss the modeling of a subsumption hierarchy consisting of UML domain model classes, caDSR object classes, and NCI Thesaurus concepts, which results in the presence of inconsistencies within the corresponding OWL ontologies. We argue that these inconsistencies are already present within the semantics of caBIG®, and we describe the way in which our querying methods take into account these inconsistencies. Second, we discuss the modeling of UML attributes as datatype properties and caDSR associations as object properties, and their relationship with associations in NCI Thesaurus, exploring the utility of these associations when constructing and executing queries.

AUTHORS AND AFFILIATIONS: Yves R. Jean-Mary, M.S.[1]; E. Patrick Shironoshita, M.S.[1]; Ray M. Bradley[1]; Mansur R. Kabuka, Ph.D.[1,2] [1]INFOTECH Soft, Inc.; [2]University of Miami.

88. The GeneTegra Information Integration System for caGrid® Data Services

GeneTegra is a graphical information integration environment designed to facilitate concept-based search and querying of genetics and biomedical data from diverse and heterogeneous data sources. It utilizes Semantic Web technologies to address the two main obstacles in the integration of knowledge: syntactic heterogeneity, where data sources have different representation and access mechanisms, and semantic variability, where similar lexical terms may refer to multiple concepts or dissimilar terms may refer to the same concept. semCDI. a semantic representation of cancer-related data services available through caGrid and a methodology to formulate and execute queries against this representation, has been enhanced to work together with GeneTegra's mechanisms to generate ontology representations of data sources, discover mappings between these ontology representations, and perform queries against these representations, enabling the integrated querying of sources within and outside of caGrid. In this poster, we present a description of GeneTegra and of its incorporation of semCDI, and show initial results that demonstrate the validity and utility of the system.

AUTHORS AND AFFILIATIONS: Yves R. Jean-Mary, M.S.[1]; E. Patrick Shironoshita, M.S.[1]; Ray M. Bradley[1]; Mansur R. Kabuka, Ph.D.[1,2] [1]INFOTECH Soft, Inc.; [2]University of Miami.

89. Planning and Execution of Queries against caGrid Data Services

caGrid provides a standard mechanism for representing the semantics of grid-enabled cancer datasets; however, integrated querying of the datasets at a conceptual level remains a challenge. The semCDI query formulation defines a methodology to specify queries against ontology representations of caGrid data services utilizing the SPARQL Query Language for RDF. The execution of these queries within the caGrid environment requires the resolution of three important issues. First, the queries must be subdivided into sub-queries specific to each data service involved. Additionally, these SPARQL sub-queries must be converted into queries written in the CQL language used within caGrid. And further, the CQL queries must be planned and executed. In this poster, we present the design and implementation of a mechanism for query planning and execution for semCDI. Specifically, we illustrate the use of the semQA query algebra for SPARQL as the mathematical foundation for the subdivision of queries, we

discuss the planning solutions implemented to overcome the limitations of CQL when joins between multiple objects are required, and we illustrate the process of execution of queries against multiple caGrid data services as implemented within our GeneTegra information integration system.

AUTHORS AND AFFILIATIONS: Yves R. Jean-Mary, M.S.[1]; E. Patrick Shironoshita, M.S.[1]; Ray M. Bradley[1]; Mansur R. Kabuka, Ph.D.[1,2] [1]INFOTECH Soft, Inc.; [2]University of Miami.

90. Extending HL7-LOINC-NCI Terminologies for Reporting Genetic Test Results

Approximately, 70-80% of electronic health records (EHR) and clinical trail data originate from the laboratory (1, 2), and caBIG® infrastructure leverages these data for research within caGrid. Unfortunately, most DNA based clinical genetic tests are currently performed by a network of dispersed labs and reported via faxed reports. However, as the cost of testing drops rapidly and becomes a part of routine clinical care, these tests may move to centralized labs already reporting structured test results (via an HL7 interface) into electronic medical records. To support this future paradigm, the Office of the National Coordinator (ONC) has developed a use case detailing requirements for the electronic medical record, and Health Level Seven (HL7) and Logical Observation Identifiers Names and Codes LOINC have published standards for reporting structured (DNA based) clinical genetic tests from the lab into an EHR. Extending this work, HL7, LOINC, the National Library of Medicine's Lister Hill Center, the National Cancer Institute (NCI), and Clinical Data Interchange Standards Consortium (CDISC) are reviewing the terminologies for these findings to determine how NCI and others can best make this information available for translational medicine and clinical research, supporting personalized cancer and other clinical medicine. See HIPAA and the clinical laboratory: changes in data handling necessitated by HIPAA regulations point clinical labs in the direction of Internet-based "wrapper" solutions - Laboratory Systems [Internet]. 2001;[cited 2010 Jul 6] Available from: http://findarticles.com/p/articles/mi_m0DUD/is_11_22/ai_80301578/2. P. Pochon, Personal Communication, July 7, 2010.

AUTHORS AND AFFILIATIONS: Mollie Ullman-Cullere, M.S., M.S.E.[1,2]; Phil Pochon[2,3,4]; Bron Kisler[4]; Margaret Haber, B.A., B.S.N., R.N.[5]; Stan Huff, M.D[6,7]; Clement McDonald, M.D.[8]; Jomol Mathew, Ph.D.[1] [1]Clinical and Translational Informatics, Dana-Farber Cancer Institute; [2]Health Level Seven International; [3]Covance Inc.; [4]Clinical Data Interchange Standards Consortium; [5]NCI CBIIT; [6]BioMedical Informatics, University of Utah; [7]Medical Informatics, InterMountain Healthcare; [8]Lister Hill National Center for BioMedical Communications, National Library of Medicine.

91. LexEVS 6.0

LexEVS 6.0 represents a major milestone as a reference implementation of Common Terminology Services 2 (CTS 2) functionality. In this release, LexEVS has been enhanced to provide additional functionality in the areas of administration, authoring/curation, and search/query of controlled vocabularies and ontologies. These are key components that are used within caBIG® and are crucial for the semantic interoperability of clinical research information. We highlight the enhancements to LexEVS 6.0 functionality, which include: Administrative Operations; Import Code System Content; Export Code System Content; Notification of Code System Content Changes; Search and Query Operations; Search and Query of Code Systems; Search and Query of Value Sets; Search and Query of Concept Domains and Usage Concepts; Search and Query of Associations/Mappings; Authoring and Curation Operations; and Code System Content.

AUTHORS AND AFFILIATIONS: Christopher Chute, M.D., Ph.D.[1]; Craig Stancl[1]; Scott Bauer[1]; Kevin Peterson[1]; Sridhar Dwarkanath[1]; Traci St.Martin, P.M.P.[1] [1]Mayo Clinic.

92. Common Terminology Services 2 (CTS 2)

Common Terminology Services 2 (CTS 2) provides a standardized interface for the usage and management of terminologies. We will provide an overview of CTS 2 and the HL7/OMG process being followed to define the standard. An overview will be given of the thematic areas considered in CTS 2: Administration; Search and Query; Authoring; and Associations. We explain the fundamental components of CTS 2 and expand on why each is important to the caBIG® community.

AUTHORS AND AFFILIATIONS: Christopher Chute, M.D., Ph.D.[1]; Craig Stancl[1]; Harold Solbrig[1]; Russ Hamm[2]; Traci St.Martin, P.M.P.[1] [1]Mayo Clinic; [2]Apelon, Inc.

93. Harmonizing Clinical Study Data Elements in Heterogeneous Meta-Data Repositories Using the CSHARE/LexWiki Environment

The CDISC Shared Health and Research Electronic Library (CSHARE) wiki is a collaborative authoring environment that enables community members to harmonize data elements from multiple organizations. The CSHARE is based on the LexWiki platform for collaborative terminology authoring, which was deployed by the NCI as the BiomedGT wiki. In the CSHARE wiki, we developed a formal UML model for data element content representation. The UML model also described how the loaded content was mapped to terminology, data types, and the points at which the content would be aligned. We also established a harmonization process that involved into three steps: (1) Semantic annotation, i.e., adding names, definitions and semantic categorization (using BRIDG data elements, ISO 21090 data types and external data standards e.g. SNOMED CT) to the individual data elements; (2) Selecting and sorting the annotated data elements to locate those that were closely related; and (3) Creating new definition for one or more common data elements that represent the community semantics. In this presentation, we explored the mechanism on exchanging data between CSHARE wiki environment and external meta-data repositories (MDRs). We specified the mappings between CSHARE wiki model schema and the schemas of the MDRs. The MDRs, including openMDR, cgMDR, caDSR, and NCBO BioPortal, were tested in a prototype, demonstrating feasible connectivity with the MDRs through a RESTful web service interface. Leveraging the built-in semantic web technology, we also investigated the approach that provides a standard query service interface (i.e. a SPARQL endpoint) on the harmonized common data elements for external reuse. In summary, we extended the capability of the CSHARE wiki and made it as a flexible, scalable, and federated environment for the storage, curation, and harmonization of clinical study data elements, which would facilitate the standardization efforts required for the whole caBIG® clinical research community.

AUTHORS AND AFFILIATIONS: Guoqian Jiang[1]; Harold Solbrig[1]; Robert Freimuth[1]; Christopher Chute[1]. [1]Division of Biomedical Statistics and Informatics, Mayo Clinic College of Medicine.

94. LexWiki and Web-Protg Integration: A Comprehensive Framework of Collaborative Authoring for Large-scale Biomedical Terminologies

LexWiki is a collaborative effort led by Mayo Clinic for development of a collaborative authoring platform for large-scale biomedical terminologies. LexWiki currently is at the core of community-based development of Biomedical Grid Terminology (BiomedGT), which enables groups of domain experts to collaboratively develop and maintain terminologies, including BiomedGT, subsets of BiomedGT, and other standalone terminologies. Web-Protg has been developed as a web front-end for the Collaborative Protg server. The core component of the Collaborative Protg

is the integration of a Change and Annotation Ontology (CHAO) into the system and the CHAO provides the basis for annotation of ontology elements, such as classes, properties, individuals, and annotation of ontology changes, such as class creation, deletion, renaming, etc. In this presentation, we propose a comprehensive framework to support the collaborative authoring of large-scale biomedical terminologies, which comprises three modules: (1) a lightweight editor in semantic wiki machinery for structured proposal creation and consensus building; (2) a formal ontology editor for change commitment, formalism rendering, and consistency checking; and (3) a terminology service module based on the LexGrid formal terminology model. We developed a prototype of the framework based on a real-world use case (i.e. WHO ICD11 revision) through a mash-up of LexWiki and Web-Protg. We demonstrate the benefits from synergizing the strengths of both environments, and discuss the challenges of the framework and relevant workflow issues.

AUTHORS AND AFFILIATIONS: Guoqian Jiang[1]; Harold Solbrig[1]; Christopher Chute[1]. [1]Division of Biomedical Statistics and Informatics, Mayo Clinic College of Medicine.

95. Grouping of caDSR Common Data Elements Using Non-metric Clustering of Concept Identifiers

The Center for International Blood and Marrow Transplant Research (CIBMTR) is currently curating 45 existing electronic and/or paper forms, most having multiple revisions, with approximately 5,000 unique Common Data Elements (CDEs) or individual data points total. Each CDE is composed of two components: a Data Element Concept, consisting of an Object Class and Property, each with Concept codes; and a Value Domain containing an Association with Concept codes. These Concept codes are numeric identifiers corresponding to specific attributes of the CDE. We have implemented clustering techniques that use these Concept codes to determine groupings of CDEs which are similar and likely to be related data points. This analysis will serve creating an HSCT-specific database model in the future.

AUTHORS AND AFFILIATIONS: Kirt Schaper[1]; Robinette Aley[1]; Barb Kramer[1]; Douglas Rizzo, M.D.[1] [1]Center for International Blood and Marrow Transplant Research.

96. Deployment of LexEVS for Enterprise Terminology Services: Lessons Learned

Controlled vocabularies in electronic medical record systems have been used to improve advanced decision making, data mining, quality assessment, patient-oriented clinical research, and detailed clinical documentations. These vocabularies can be categorized as external reference terminologies, created by national and international organization (like LOINC and SNOMED), or locally developed terminologies usually created for intra-organizational consumption. Interoperability across all domains within enterprise information system is heavily dependent upon such vocabularies and the terminology services that maintain them. We implemented a local version of LexEVS 5.1 as a proof of concept for our terminology services that can serve both internal and external reference vocabularies. Using LexEVS UMLS batch loader, all level 0 subsets of UMLS Metathesaurus with Rich Release Format (RRF) were imported into the LexEVS server with Oracle 11g backend. The amount of time required for import process was dependant to the vocabulary size and format (few hours for SNOMED in RRF and OWL formats, and few minutes for CTCAE in OWL format). We also used Protg 3.4.4 and TopBraid Composer to convert our locally developed terms, in excel formats, into OWL using Semantic Web standards. Vocabulary terms (including one third-party vendor terminology) were subsequently categorized as instances of classes, super classes, and

sub classes and converted using owl:Class and rdfs:subClassOf resources. Associated metadata (e.g., labels, codes, definitions) were created by rdfs:label resource and annotation data type properties. The resultant OWL file was successfully loaded into LexEVS 5.1 using included LexEVS OWL loader. We also implemented a local version of publicly available NCI Term browser for visualizing imported vocabularies (internal or external) and creating linkage among them with the help of NCIT Browser xml property file. Current challenges include importing third-party vocabularies in non-standard formats, handling value sets, and importing large-size standard vocabularies in OWL format.

AUTHORS AND AFFILIATIONS: Sina Madani, M.D., M.S.[1]; Jeff Jin, Ph.D.[1]; Michael Riben, M.D.[1]
[1]University of Texs, MD Anderson Cancer Center.

97. Collaborative Development of Ontologies using WebProtg and BioPortal

Development of ontologies requires an integrated platform that not only allows for ontology developers to collaboratively edit the ontology, but also allows for the collection of comments from subject matter experts. The integration of WebProtg and BioPortal provides such an environment. Web Protg is a lightweight web-based ontology browser and editor, which provides a collaborative environment for editing. These features include the ability to simultaneously browse and edit the ontology, to discuss entities in the ontology (e.g., class, property, or individual), and to track all changes to the ontology. Once the cycle of edits is complete and a new version of the ontology is generated, many ontology developers choose to release their ontology by placing it on the Web for others to access the ontology and to collect comments from the community. BioPortal provides such a mechanism to publish ontologies and collect community feedback, in addition to a range of other functionality. Using the BioPortal web services, the ontology and its metadata can be published directly to BioPortal. Subject matter experts can login to BioPortal, browse the ontology and add comments to entities and propose changes. The notes are stored in a common representation and therefore, the ontology editors can see the note in the context of editing the ontology and make the needed changes. Notes can be archived once the editing task is complete, which will then hide the note from view in BioPortal. Requirements for the implementation of an integrated editing platform for ontology development and publishing have been collected from review of existing tools and workflows of large collaborative ontology development projects. One of the main drivers and users of this collaborative platform is the World Health Organization in development of ICD-11.

AUTHORS AND AFFILIATIONS: Patricia L. Whetzel, M.S., Ph.D.[1]; Natasha Noy, Ph.D.[1]; Tania Tudorache, Ph.D.[1]; Csongor Nyulas, M.S.[1]; Mark Musen, M.D.,Ph.D.[1] [1]Stanford University.

98. NCI EVS Terminology Browsers: Serving and Connecting a Growing Community

NCI Enterprise Vocabulary Services (EVS) creates, hosts, and cross-maps controlled terminology for a growing community of NCI, caBIG®, and other partners. EVS serves the requirements of researchers, clinicians, patients, and other users, while promoting shared standards and harmonization. EVS browsers are currently being redesigned and extended to support an even broader range of users, content, and services, including support for new value set and mapping services being developed jointly with LexEVS 6.0 using the HL7 Common Terminology Services 2 (CTS 2) specification. NCI Thesaurus (NCIt) controlled terminology plays a central role in the NCI/caBIG® semantic infrastructure and information services. NCIt has over 80,000 concepts, with 200,000 inter-concept relationships and extensive synonymy, definitions, and other information. Further,

NCIt is being used for collaborative terminology development by NIH, and other partner organizations such as the Food and Drug Administration (FDA), the Clinical Data Interchange Standards Consortium (CDISC), and the National Council for Prescription Drug Programs (NCPDP), who together use 148 NCIt terminology subsets with 25,000 terms. The NCIt Browser is being enhanced to help users find and interpret the content they need, especially where multiple coding standards overlap; subsets and cross-mappings will be directly surfaced in the new browser interfaces. The NCI Term Browser provides access to all biomedical terminologies hosted by EVS. This browser extends the NCIt Browser to more than 20 additional terminologies. It also supports search and cross-links to the NCI Metathesaurus (NCIm), which itself offers 76 terminologies in a convergent terminology environment, cross-mapped to capture shared meanings. The new CTS 2 services will allow storing and browsing of separately defined value sets, as well as mappings between terminologies and value sets that may be derived from NCIt, NCIm, or other sources. NCI EVS is actively exploring other improvements to its content and services, and invites user feedback and suggestions through numerous links on its browsers, web pages, and external partner sites.

AUTHORS AND AFFILIATIONS: Lawrence Wright, M.A.[1]; Wilberto Garcia, M.S.[2]; Margaret Haber, B.S.N., R.N.[1]; Steven Hunter[3]; Jason Lucas, B.S.[3]; Kim Ong, Ph.D.[2]; Tracy Safran[4]; Robert Wynne[5]; David Yee, M.S.[2] [1]NCI CBIIT; [2]Northrop Grummon; [3]Ekagra; [4]SAIC-Frederick; [5]Lockheed Martin.

99. Sharing Behavioral Measure Data using caGrid

The Population Sciences SIG aims to integrate the various areas of population sciences within caBIG® to help understand how population science data and methods can use grid technology to further the science. One of these activities involves incorporating measures frequently used by population scientists into the cancer Data Standards Registry and Repository (caDSR) in order to enable access to the datasets containing those measures on the caGrid. The Grid Enabled Measures (GEM) project, funded by the Division of Cancer Control and Population Sciences (DCCPS), is designed to promote the use of standardized measures in research and sharing harmonized data and it provides a repository for metadata about commonly used behavioral measures and their associated constructs. The ultimate goal is to allow data capture using the measures in GEM and enable sharing of this data across the caGrid based on the Common Data Elements (CDEs) this group is registering in the caDSR. The process of curating CDEs for the GEM measures into the caDSR includes UML modeling, semantic integration, and registration of metadata. UML models have been developed for several of the measures already contained in GEM. The best practices for representing measures using UML models have been investigated. One challenge of curating measures is the lack of ontology for the measures and population science in general. Currently, the Enterprise Vocabulary Services (EVS) does not contain ontologies for behavioral measures. Concepts from different sections of the NCI Thesaurus hierarchy that involve multiple terminologies are currently being used to annotate the behavioral related UML models. The Population Sciences SIG needs to work closely with the EVS team to ensure that terms related to "Behavioral Oncology" are incorporated into the NCI Thesaurus.

AUTHORS AND AFFILIATIONS: Hua Min[1]; Mike Collins[1]; Jessica Bondy[2]; Nancy Avis[3]; Paul Courtney[4]; Richard P. Moser[5]; Mary Cooper[6]; Dianne Reeves[5]; Bob Lanese[7]; Cindy Helba[8]; Riki Ohira[9]; Richa Gandhi[9]. [1]Fox Chase Cancer Center; [2]University of Colorado Denver; [3]Comprehensive Cancer Center of Wake Forest University; [4]SAIC-Frederick; [5]NCI CBIIT; [6]SAIC; [7]Ireland Cancer Center, Case Western Reserve University; [8]Westat; [9]Booz Allen Hamilton.

100. Life Sciences Subject Matter Experts Model Business Processes and Capabilities to Support ECCF

The caBIG® Life Sciences Business Architecture Model 1.1 (LS BAM 1.1) is a foundational tool supporting caBIG®s goal of providing an informatics framework for "bench to bedside and back" research. BAMs supply a common set of terminology, use cases, and actors for a domain, thus providing clear, concise descriptions of business capabilities, governance structures, business processes, and business knowledge. The LS BAM is a common reference point to communicate how Life Sciences (LS) research is performed. It can be used for tracing components of conceptual, logical, and implementation models to the standard business processes and capabilities and to inform high level requirements for constructing software and services that use LS research data, model LS research practices, or provide services for LS researchers. Working iteratively with a Business Analyst, the LS Subject Matter Experts developed the LS BAM, describing goals that are common to all LS research sub-domains, including: how LS research is planned and performed; the organizations, resources, and methodologies involved; how data and materials are controlled and analyzed; and how outputs are disseminated. The primary output from the LS BAM effort has been a Use Case UML model (for business capabilities), with 90 use cases providing significant documentation about LS business goals through detailed storyboards, pre-conditions, the basic flow of events to accomplish the goal, and post-conditions. Activity Diagrams (for business processes) graphically illustrate the chronological or logical arrangement of these goals. Since Activity Diagrams tend to document very specific processes, they are created as areas of priority are identified, and a single exemplar was constructed. Lastly, Actors are associated with use cases in Use Case and Activity Diagrams to convey ownership and responsibility. The LS BAM currently contains 61 actors, each representing a person, group, or institution.

AUTHORS AND AFFILIATIONS: Lauren Becnel Boyd, Ph.D.[1]; Scott P. Hunicke-Smith, Ph.D.[2]; Grace A. Stafford, Ph.D.[3]; Elaine T. Freund, Ph.D.[4]; Michele Ehlman[5]; Uma Chandran, Ph.D.[6]; Robert Dennis, Ph.D.[7]; Anna T. Fernandez, Ph.D.[8]; Stephen Goldstein[9]; David Steffen, Ph.D.[10]; Benjamin Tycko, M.D., Ph.D.[11]; Juli Klemm, Ph.D.[12] [1]Baylor College of Medicine; [2]University of Texas, Austin; [3]The Jackson Laboratory; [4]3rd Millennium; [5]Essex Management; [6]University of Pittsburgh; [7]University of California, Los Angeles; [8]Booz Allen Hamilton; [9]Sapient Government Services; [10]Baylor College of Medicine; [11]Columbia University; [12] NCI CBIIT.

101. xService

We present the xService infrastructure that provides support for building caGrid® data services from XML schemas and XML database backends. Generally a Data Service shares to the grid data that is stored in a relational database backend. caGrid data services for relational database backends are supported by caCore SDK and caGrid caCore service plug-ins. The xService infrastructure has been developed in response to the requirements associated with managing, exposing, and sharing datasets that conform to existing XML schemas like HL7AECG and PAIS, and was driven mainly by use cases in the CardioVascular Research Grid. XML-base caGrid Data Services can be created using the xService extension for the Introduce toolkit. However, to ensure full caBIG® compatibility of these services and their data models, object-oriented models served by such services should be semantically annotated and registered in the environment. In this poster, we will present the core xService infrastructure and describe the tools we have developed to create UML models from XML schemas, to convert these UML models to domain models used by the Introduce toolkit for service creation, and to map CQL queries against the UML models to XPath queries to

backend XML schemas and databases. By creating UML models from XML schemas, we are able to leverage the caBIG® model curation and annotation tools such as SIW. Our UML generation workflow adds special tags to a UML model (more specifically to the corresponding XMI document) so that the mapping between the UML model and the XML schema, from which the UML model is generated, is maintained. This is necessary as the UML model may be modified during the caBIG® semantic annotation and harmonization process. These special tags are also used when generating domain models for caGrid data services and for correctly mapping CQL queries to backend XPath queries.

AUTHORS AND AFFILIATIONS: Tony Pan[1]; Fusheng Wang[1]; Tahsin Kurc[1]; Cristobal Vergara Niedermayr[1]; Ashish Sharma[1]; Joel Saltz[1]. [1]Emory University.

102. Information Representation Working Group Information Model Development To Support ECCF

The NCI Center for Biomedical Informatics and Information Technology (NCI CBIIT) is implementing a Service Oriented Architecture using the HL7 Services-Aware Interoperability Framework (SAIF). The Enterprise Conformance and Compliance Framework (ECCF) is a subframework of SAIF and is a tactical approach to describing service specification and enabling interoperability. The Information Representation Working Group (IRWG), a component of the Integrative Cancer Research Workspace in the caBIG® program, is focused on developing information models to be used to support ECCF and the development of interoperable Life Sciences (LS) applications. These models provide common semantics that facilitate traceability between business requirements, technical requirements and service specifications.The LS Domain Analysis Model (LS DAM) is an object-oriented information model that provides a high-level common representation of the most relevant types of data generated in translational research. It is an implementation-independent view of the static semantics of the domain and includes classes, attributes and the associations among those classes.The LS Reference Platform Independent Model (PIM) is a constrained and localized model of subdomains of the LS DAM that is not dependent on the development or deployment environment or specific technologies. This model utilizes core DAM concepts and provides direct traceability to the LS DAM. It is intended as a starting point for project information modeling and has already been used by the caBIO team for the development of the Molecular Annotation Service. Future IRWG activities include expansion of the LS DAM, further alignment of LS DAM to BRIDG, mapping to HL7 RIM, adding the NCI adopted standard of ISO21090 datatypes and collaborating with other standards development organizations. These will further enhance the utility of the LS DAM and LS PIM as they are used to support development of LS applications and services that enable information integration across the bench to bedside continuum.

AUTHORS AND AFFILIATIONS: Robert Freimuth, Ph.D.[1]; Elaine Freund, Ph.D.[2]; Jenny Kelley[3]; Juli Klemm, Ph.D.[4]; Sue Pan, Ph.D.[5]; Konrad Rokicki, M.S.[5]; Lisa Schick[6]; Mukesh Sharma, Ph.D.[7]; Grace Stafford, Ph.D.[8]; Baris Suzek, M.S.[9]; Dennis Thomas, Ph.D.[7] [1]Mayo Clinic; [2]3rd Millennium; [3]NCI LPG; [4]NCI CBIIT; [5]SAIC; [6]ScenPro; [7]Washington University in St. Louis; [8]The Jackson Laboratory; [9]Georgetown University.

103. caAdapter Model Mapping Service:Tooling Support for ISO 21090 Development

caAdapter Model Mapping Service (caAdapter MMS) is part of the integrated development tools to support NCI CBIIT's localization of ISO 21090 data types and service development. As part of the caAdapter tool set, caAdapter MMS leverages the caAdapter's mapping infrastructure to map an object model to a

data model. The mapped models can then be used by caCore SDK to build application systems. caAdapter MMS supports both standalone and web deployments.

AUTHORS AND AFFILIATIONS: S. Liu[1]; C. Andonyadis[1]; A. Basu[1]; D. Warzel[1]; A. Shanbhag[1]; E. Wang[2]; K. Um[2]; Y. Wu[2]; W. Ver Hoef[3]; S. Hastak[3]. [1]NCI CBIIT; [2]SAIC; [3]ScenPro, Inc.

104. caAdapter Data Mapping and Transformation Service: A Bridge to caBIG® Interoperability

caAdapter Data Mapping and Transformation Service (caAdapter CMTS) is part of the caAdapter tool set. It supports mapping and transformation among different kinds of data sources including CSV, XML, and HL7 messages. It also provides a set of data manipulation functions for data transformation. caAdapter CMTS supports NCI CBIIT's localization of ISO 21090 data types. It provides web service interfaces for SOA integration, as well as Java APIs for application integrations.

AUTHORS AND AFFILIATIONS: A. Basu[1]; D. Warzel[1]; A. Shanbhag[1]; E. Wang[2]; K. Um[2]; Y. Wu[2]; W. Ver Hoef[3]; S. Hastak[3]. [1]NCI CBIIT; [2]SAIC; [2]ScenPro, Inc.

105. LS DAM: A Foundation for Building Semantically Interoperable Services In The Life Sciences Domain

As NCI CBIIT moves further into the semantic service-oriented architecture with the adoption of the HL7 Service Aware Interoperability Framework (SAIF), the requirement for sharing data and supporting working interoperability across the various functional domains will be the key to success. One of the major components of the SAIF's Information Framework is Domain Analysis Models (DAMs). These DAMs provide the foundation for the Information Viewpoint in RM-ODP and the static semantics for building the Semantic Profiles in the ECCF service specifications.The Life Sciences Domain Analysis Model (LS DAM) provides the computable representation of information semantics for the collective domains of Integrative Cancer Research (ICR), Tissue Banking and Pathology Tools (TBPT), and Imaging. The LS DAM continues to evolve since its initial release in July 2009. Release 1.2 (February 2010) was aimed at providing a framework of support for the key components of the ICR Nano sub-domain. The LS DAM will continue to evolve to more completely support the caBIG® Life Sciences workspaces, driven and governed by an analyst, the ICR Information Representation Working Group (IRWG), and other subject matter experts as identified by the Life Sciences Governance Team (LGT). Plans for future releases of LS DAM include: alignment with BRIDG 3.0; Introduction of a multi-layered approach, where each layer is directed at supporting the needs of a specific stakeholder group (a domain friendly upper layer; a comprehensive middle layer; and an HL7 RIM-based lower layer); inclusion of more comprehensive support for the Person and Organization roles within the Life Sciences domain; and extension to provide more complete support for key concepts, including: Experimental Design and Outcomes, Pathway/Molecular Interactions, Biomarkers, the Nano sub-domain, and specimen distribution and handling.

AUTHORS AND AFFILIATIONS: Robert R. Freimuth, Ph.D.[1]; Elaine Freund, Ph.D.[2]; Sharon Gaheen[3]; Jenny Kelley[4]; Juli Klemm, Ph.D.[5]; Sue Pan, Ph.D.[3]; Konrad Rokicki, M.S.[3]; Lisa Schick[6]; Mukesh Sharma, Ph.D.[7]; Grace Stafford, Ph.D.[8]; Baris Suzek, M.S.[9]; Dennis Thomas, Ph.D.[7] [1]Mayo Clinic; [2]3rd Millennium; [3]SAIC; [4]NCI LPG; [5]NCI CBIIT; [6]ScenPro; [7]Washington University in St. Louis; [8]The Jackson Laboratory; [9]Georgetown University.

106. caGrid Portal

The caGrid Portal is a web-based application that enables you to discover and interact with the services that are available on the caGrid infrastructure. The portal serves as the primary visualization tool for

the caGrid middleware and provides a standards-based platform for hosting caBIG®-related tools. It also serves as a caBIG® information source. It provides access to information about caBIG® participants, points of contact (POCs), and caGrid-related news and events. The latest release supports sharing of content like caBIG® tools, services, and caGrid data service queries with other users of the caBIG® community, and the creation of ad hoc communities that can be administrated by portal users. Communities are fully customizable web spaces within the portal that are provisioned with common collaboration tools (e.g. wiki, forums, document library, WYSIWYG content editors, etc.) to facilitate sharing of data and knowledge around specific areas of interest.

AUTHORS AND AFFILIATIONS: Joshua Phillips[1]; Manav Kher[1]. [1]SemanticBits.

107. QA Initiatives at NCI CBIIT

The QA team at NCI CBIIT has implemented several new Standard Operating Procedures (SOPs) and tools that have increased traceability, transparency, and the overall quality of CBIIT products. By mapping requirements to test cases in a Requirement Traceability Matrix (RTM), we are able to easily measure the impact of a defect on functionality and end users. Transparency is achieved using shared documentation within the project team. This leads to the increased communication of status and results of QA's work to the rest of the members of the project. As a result, stakeholders are able to track the results of test execution in real time using a dashboard tool. These new initiatives and tools increase the overall product quality by streamlining lines of communication between different team members and removing ambiguity of the functional coverage.

AUTHORS AND AFFILIATIONS: Steve Hunter[1]; Cliff Wong[1]; Sudha Chudamani[1]; Ye Wu[1]; Quy Phung[1]. [1]NCI CBIIT.

108. Enterprise Service Specifications Team (ESST)

The Enterprise Service Specifications Team (ESST) is constituted by NCI's CBIIT to help govern the specification process of Enterprise Compliance and Conformance Framework (ECCF). ESST's primary role is to administer and manage the lifecycle of ECCF-compliant Service Specifications for NCI's Enterprise Services (NES). ESST assists the Governance Team in identifying new services and determining their scope as well as in the management of the NES portfolio. ESST has also created various guidelines, standards, and review criteria to ensure accuracy and consistency in NES specifications. Once a service development team has been assigned, ESST assists them in creation of the specifications by providing guidance and support. ESST also conducts a preliminary review of the specifications before presenting them to the respective Composite Architecture Teams (CATs) for approval. As analysts, ESST ensures that the services as specified will meet the overall business objective and provide all the capabilities needed by a particular business domain. As architects, ESST ensures that the services adhere to the standards and guidelines set by NCI's enterprise architecture. Lastly, the team also helps ensure that the services are interoperable - both syntactically and semantically.

AUTHORS AND AFFILIATIONS: Brian McIndoe[1]; Kunal Modi[1]; Santosh Joshi[1]; Kalpesh Patel[1]; Jeff McLean[2]; Christo Andonyadis[3]. [1]Ekagra Software Technologies; [2]SAIC-Frederick; [3]NCI CBIIT.

109. The Laboratory Information Digital Data Exchange (LIDDEx) Consortium:Two years of experience with the development of a semantically-encoded interoperability layer for clinical and investigative laboratory results

Although the use of information technology and database abstraction layers has been firmly in place within the pathology

subspecialty practices of clinical laboratory medicine and surgical pathology for over 40 years, there remains a continuing deficit of functionality in the specific area of inter-institutional and extramural exchange of laboratory results due to limitations including: a continued lack of a standard messaging construct (recognizing that HL7 2.X is not such a standard) and lack of a critical mass of consortial partnership, on the part of both the laboratory information vendor space and the end-user pathologist-laboratorian space, to enable the creation of a single, unified standard. The LIDDEx consortial initiative represents a realization of these two enabling conditions, and demonstrates that interoperability is indeed attainable and enabled by collective participation and development on the part of the current consortial partners. In the current third-generation implementation exercise, the LIDDEx consortium is well positioned to demonstrate the utility of a fully semantically-encoded data interchange channel across a grid of disparate federants, in real time, with the added capability of a University of Michigan Division of Pathology Informatics' pilot enterprise master-patient index (EMPI)-based identity adjudication (also in real time) with the added feature of state-based buffering of demographic updates, in support of synchronizing federant nodes that may be asynchronously detached at the time of original demographic construct update. Collectively, the LIDDEx consortial solution represents a promising exemplar architecture for longitudinal and geographically dispersed aggregation of results from multiple repositories, as are typically encountered in the caGrid continuum. Given this commonality, and the current prevailing design directives of the LIDDEx consortial effort to ultimately adopt or transition to caBIG®-compliant schemata and service classes whenever possible, it is hoped that such commonality will facilitate the ultimate integration of LIDDEx functionality, and thus lab results interoperability, into the extant repertoire of caBIG® service classes.

AUTHORS AND AFFILIATIONS: Ulysses J. Balis[1]; Jeff Sica[1]; John Hamilton[1]. [1]Division of Pathology Informatics, University of Michigan Health System.

110. Leveraging caGrid Technologies for Linking Clinical and Biospecimen Data Repositories

As part of the clinical and translational research efforts at The Ohio State University, investigators at OSUCCC (The Ohio State University Comprehensive Cancer Center) wish to share and integrate data located in distributed repositories. The overall goals of the caGrid project[1] align with this specific need and consequently we leverage caGrid technologies to fulfill the data needs of scientists and investigators at OSUCCC. As a primary use case, we linked data in the OSUMC information warehouse (clinical data repository) with that in participating tissue banks (bio-specimen repository). Moreover, we also setup a Federated Query Processor (available as part of caGrid distribution) that can handle queries with specific constraints on the involved data sources. This provides us with the ability to provide filtered and highly relevant data to an investigator's request. The tissue banks use caBIG®'s caTissue suite[2] for bio-specimen management. The caTissue suite has a data service that is registered onto a grid (the TRIAD grid in our case). Similarly, we have deployed a data service that can communicate with the information warehouse datamart (specific view of the information warehouse that can be accessible via the grid). Finally, the front end is provided by the caGrid portal[1] deployment containing a custom portlet that captures user needs in terms of query constraints and translates them to DCQL queries that are in turn run on a Federated Query Processor. References:1: Oster, S. et al., (2007), "caGrid 1.0: An Enterprise Grid Infrastructure for Biomedical Research", Journal of the American Medical Informatics Association, Vol. 15(2): pp.138-149.; 2: Klemm, J. et al. (2010), The caBIG® Life Sciences Distribution, In Ochs, M.F.,

Casagrande, J.T., Davuluri, R.V. (eds.), Biomedical Informatics for Cancer Research, 1st edition, Springer, pp. 253-266.

AUTHORS AND AFFILIATIONS: Omkar Lele, M.S.[1]; David Ervin[1]; Justin Permar[1]; Tara Borlawsky, M.A., Philip R. O. Payne, Ph.D.[1] [1]Center for IT Innovations in Healthcare, Department of Biomedical Informatics, The Ohio State University.

111. Triton: An Integrative Translational Research Information Management Platform

The Chronic Lymphocytic Leukemia (CLL) Research Consortium (CRC), as a prototypical instance of a multi-site research consortium, uses numerous loosely coupled web applications to address its informatics needs. More than a decade of operations have allowed the CRC to identify usability and computational limitations relative to the preceding information management architecture. In response, the CRC has launched the TRITON (Translational Research Information Technology OmNibus) project, with the ultimate objective of developing an open-source, extensible, and fully integrative translational research information management platform. A primary goal of these efforts is to enable the integration between TRITON and basic science, clinical research, and translational science-focused data management tools and interchange mediums associated with the NCI's Cancer Biomedical Informatics Grid® (caBIG®) initiative, including the caGrid service-oriented middleware, caTissue, and caAERS. In doing so, our objectives is to increase the translational capacity of the CRC by enabling consortium investigators to discover, integrate, analyze, and disseminate heterogeneous, multi-dimensional data sets.This poster will provide an introduction to TRITON including an list of participating CRC institutions, a description of the development effort between OSU and UCSD, system architecture, description of caBIG® technology adoption (caGrid, caTissue, caAERS), a description of caBIG® technology adaptation (a custom caTissue grid service), current status, and future plans. References: 1: Oster, S. et al., (2007), "caGrid 1.0: An Enterprise Grid Infrastructure for Biomedical Research," Journal of the American Medical Informatics Association, Vol. 15(2): pp.138-149.;2: Klemm, J. et al. (2010), The caBIG® Life Sciences Distribution, In Ochs, M.F., Casagrande, J.T., Davuluri, R.V. (eds.), Biomedical Informatics for Cancer Research, 1st edition, Springer, pp. 203-213, 253-266.

AUTHORS AND AFFILIATIONS: William Stephens[1]; Andrew Greaves[2]; Justin Permar[1]; Philip R. O. Payne, Ph.D.[1] [1]Center for IT Innovation in Healthcare, Department of Biomedical Informatics, The Ohio State University; [2]Moores Cancer Center, University of California San Diego.

112. Using the Open Metadata Registry (OpenMDR) to Generate Semantically Annotated Grid Services

OpenMDR is a metadata registry capable of storing, versioning, and maintaining semantic and representational metadata. It provides a suite of tools consisting of four different components: (1) MDR Core; (2) MDR Query; (3) MDR Plug-in; and (4) MDR Domain Model Generator. MDR Core is an ISO11179 semantic repository. MDR Query is a grid service used to search multiple semantic repositories (including MDR Core). UML modelers use MDR Plug-in to search for semantic concepts from within Enterprise Architect. caGrid Service Developers use the MDR Domain Model Generator to create caGrid data services. Each of these projects provides functionality that enables federated semantic metadata annotations to be created and used in Grid Service Registration and Discovery. This will now give caGrid service developers the option of using other semantic metadata management tools in addition to that of the NCI caDSR, EVS, and others. Additionally, developers will maintain a fast and agile process for annotating and delivering a

strongly typed and semantically anchored grid service into production that is independent from policies concerning the use of particular metadata registries. References include: (1) Oster, S. et al.,"caGrid 1.0: A Grid Enterprise Architecture for Cancer Research", Proceedings of the 2007 AMIA Annual Symposium: pp.573-577, December 2007; and (2) Cancergrid's cgMDR, UK (http://cancergrid.org).

AUTHORS AND AFFILIATIONS: Rakesh Dhaval, M.S.[1]; Calixto Melean, M.S.[1]; David Ervin[1]; Philip R. O. Payne, Ph.D.[1] [1]Center for IT Innovation in Healthcare, Department of Biomedical Informatics, The Ohio State University.

113. Adaptation and Adoption of caBIG® Tools to a Clinical Research Community

The Ohio State University Clinical and Translational Science Award (CTSA) is addressing the need to collect, store, integrate, analyze, and report upon heterogeneous and distributed data sets spanning organizational boundaries. This is accomplished by a grid middleware and tooling system known as TRIAD. TRIAD addresses the needs and challenges presented by multi-institutional translational team science by allowing for the creation of a scalable, secure, and knowledge anchored data sharing environment. TRIAD builds upon and extends the robust caGrid infrastructure that was initially designed for the National Cancer Institute's caBIG® program.TRIAD seeks to integrate disparate clinical data sources into a common grid environment. By doing so, researchers maintain locus of control over their research while expanding the scope, capability, and quality of research and knowledge made available to the broader TRIAD adopter community. TRIAD accelerates the pace of research by providing the means to create knowledge centered pipelines and facilitating cross-institutional data queries. Specialized adapters have been created within TRIAD to enable use of common data gathering and storage tools such as REDCap[1] and i2b2[2] within the existing grid infrastructure. By approaching data sharing from the perspective of a wrapper around an existing tool rather than a monolithic and large engineering effort, researchers and clinicians may readily and rapidly move to interconnected grid systems. References include: (1) Paul A. Harris, Robert Taylor, Robert Thielke, Jonathon Payne, Nathaniel Gonzalez, Jose G. Conde, Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support, J Biomed Inform. 2009 Apr;42(2):377-81; and (2) Shawn N. Murphy, Griffin Weber, Michael Mendis, Vivian Gainer, Henry C. Chueh, Susanne Churchill, Isaac Kohane, Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2), JAMIA 2010;17:124-130 doi:10.1136/jamia.2009.000893.

AUTHORS AND AFFILIATIONS: David Ervin[1]; Sreekant Lalokta[1]; Puneet Mathur[1]; Rakesh Dhaval, M.S.[1]; Calixto Melean, M.S.[1]; Omkar Lele, M.S.[1]; Tara Payne, M.A.[1]; Philip R. O. Payne, Ph.D.[1] [1]Center for IT Innovation in Healthcare, Department of Biomedical Informatics, The Ohio State University.

114. caGrid 1.4

The caGrid middleware is a service-oriented platform that forms the backbone of the grid infrastructure employed by caBIG® to facilitate secure, multi-institutional e-science collaborations. Such collaboration is possible because caGrid provides an interoperable service-oriented architecture to share data and analytical resources. caGrid uses metadata to describe the structure and semantics of resources and also provides discovery capabilities to find resources of interest. The 1.4 release of caGrid is the next evolutionary step of the caGrid® platform which began with the initial 1.0 release in 2006. caGrid 1.4 adds new features driven by feedback from the user community, caGrid adopters, and CBIIT technical leadership. As caGrid provides the

supporting framework for many NCI and CBIIT projects, it has evolved to incorporate the requirements of a large and diverse community. This support includes such items as support for ISO 21090 data types, integration of the out-of-band high throughput data transfer framework in Federated Query Processing, and support for the latest versions of CBIIT's data system development toolkit, the caCore SDK.

AUTHORS AND AFFILIATIONS: David Ervin[1]; Calixto Melean, M.S.[1]; Justin Permar[1]; Joe George[1]; Albert M. Lai, Ph.D.[2] [1]Center for IT Innovations in Healthcare, Department of Biomedical Informatics, The Ohio State University; [2]Department of Biomedical Informatics, The Ohio State University.

115. CTRP Trial Registration Service: NCI's ECCF Compliant Enterprise Service

The fundamental goal of caBIG® Enterprise Compliance is the creation and adoption of software systems, tools, and services that are syntactically and semantically interoperable and fulfill the objectives of building an infrastructure for achieving predictable Working Interoperability (WI). These guidelines define processes for supporting computable semantic interoperability among the various development efforts in caBIG®. The CTRP/COPPA project is one of NCI's reference implementations of building a SAIF-based Enterprise Architecture. CTRP/COPPA is bound to the Enterprise Architecture Specifications (EAS) of BRIDG as the Domain Analysis Model (DAM), HL7 RIM, ISO 21090 data types and the standardized vocabulary bindings of clinical research. The EAS also supports a rigorous governance process and adoption of the Enterprise Conformance and Compliance Framework (ECCF). The ECCF defines the various layers of abstraction and representations of services that are needed to provide the required levels of detail for the different project stakeholders. CTRP's Trial Registration and Management Service is the first ECCF-compliant business service in NCI's Enterprise Services (NES) portfolio. The ScenPro development team that developed this CTRP/COPPA service also documented the analytical service using the ECCF specifications for the CIM, PIM and PSM. The Trial Registration and Management business service provides a standardized set of interfaces for creating, updating, and amending proprietary and non-proprietary interventional trials within the NCI's Clinical Trials Reporting Program (CTRP) system. The business service internally executes various granular COPPA RIM services such as Entity, Role, Participation, and Act that are designed and represented in the BRIDG-based CTRP Information model. This service provides vendors, cancer centers, and other consumers of this service the capability to build service-level integration of their systems with CTRP services in order to register, update, and amend their trials by consuming domain-oriented service interfaces that hides the complexity of the RIM-based COPPA/CTRP model.

AUTHORS AND AFFILIATIONS: Naveen Amiruddin[1]; Hugh Reinhart[1]; Nellie Martchenko-Shimko[1]; Smita Hastak,[1]; Kalpana Guthikonda[1]; Anu Sharma[1]; Christo Andonyadis[2]; John Speakman[2]; Steve Silberman[3]. [1]ScenPro, Inc.; [2]NCI CBIIT; [3]SAIC-Frederick.

116. caBIG® Interoperability Scenarios

The caBIG® enterprise architecture has matured over the last 5 years and the focus now is to support cross-workspace interoperability across all the domains of healthcare. To truly achieve the goal of semantic working interoperability, the caBIG® Enterprise Interoperability Scenarios Team has been charged with developing enterprise interoperability use cases that identify pain points in the process of executing clinical trials. The team is focused on building scenarios that will support working interoperability between the traditional boundaries of hospital systems and research systems. One of the main areas of interest for the team is to identify

and define interoperability scenarios that cross into the caBIG® workspaces, Imaging, Integrative Cancer Research (ICR), Tissue Banking and Pathology Tools (TBPT), and the hospital systems. The team has extensively defined the detailed interactions of 10 scenarios in the form of UML Activity diagrams. The static data elements exchanged during these scenarios have been harmonized with the BRIDG analysis model. The team has built the dynamic and the static structures needed to support the implementation of the scenarios. The team will be working with the various project teams in implementing a select group of scenarios and identifying the common services for the caBIG® Enterprise infrastructure. These interoperability scenarios have also been harmonized with the CTMS Business Architecture Model (BAM) and map to multiple BAM use cases.

AUTHORS AND AFFILIATIONS: Vijaya Chadaram, R.N., B.S.N.[1]; Sharon Elcombe, M.A.[2]; Jomol Mathew, Ph.D.[3]; Mary Agnes Templeton, CCRP[3]; Wendy Ver Hoef[4]; Smita Hastak, M.S.[4] [1]Duke University Medical Center; [2]Mayo Clinic; [3]Dana Farber Cancer Center; [4]ScenPro, Inc.

117. The Cancer Therapy Evaluation Program Enterprise System (CTEP-ESYS) Service-Oriented Architecture (SOA)

Service-Oriented Architecture (SOA) is an approach for organizing IT resources in which data, logic, and infrastructure resources are accessed by routing messages between networked interfaces. SOA enhances the interoperability and data sharing across heterogeneous systems through the creation of well-defined, self-sustained services that can intercommunicate via standard message interfaces. The Cancer Therapy Evaluation Program Enterprise System (CTEP-ESYS) is a collection of heterogeneous systems designed to enhance the scientific and administrative aspects of cancer clinical trial development. The CTEP-ESYS was developed by the Cancer Therapy Evaluation Program (CTEP), which oversees the research conducted investigating new distinctive and effective anticancer agents, radiation treatments, and surgical methods. CTEP is now focusing on integrating the CTEP-ESYS systems to improve the data exchange and workflow across trial phases and to support better interoperability. This integration includes migrating the CTEP-ESYS to a SOA and to leverage the existing technology base through exploitation of the best components developed so far. The Online Agent Order Processing application (OAOP) will be the first application to use the CTEP-ESYS SOA framework and will integrate with FedEx Web services. The Adverse Event Reporting System (AdEERS) web services will fully harmonize with caBIG®'s cancer Adverse Event Reporting System (caAERS) for data exchange and adverse event reporting with CTEP-ESYS. CTEP Enterprise services will fully integrate with NCI Enterprise Services (NES) for data exchange with CBIIT, a BRIDG 2.0 compliant model using ISO 21090 data types. SOA will enhance the clinical trial development process and collaboration within the research community by converging with the Cancer Biomedical Informatics Grid® (caBIG®). It will also improve the interoperability and data sharing across heterogeneous systems, the connections between scientists and practitioners through a shareable and interoperable infrastructure, and facilitate the implementation of standard rules and a common language as defined by caBIG® for data sharing.

AUTHORS AND AFFILIATIONS: Steven Friedman, M.H.S.A.[1]; George Redmond, C.G.E.I.T., M.Sc., M.B.A.[1]; Shanda Finnigan, R.N., B.S.N, C.C.R.C.[1]; Yogi Byreddy, B.S., P.M.P.[2]; Crystal Evans, B.A., P.M.P.[2]; Hariprasad Davanagere, M.S.[2]; Ganesh Panathula, B.E.[2] [1]NCI, DCTD, CTEP; [2]CTIS, Inc.

118. Revisiting Address Interoperability in a SAIF Environment

A small working group used the Address use case to understand how adoption of the ISO 21090 datatype standard and domain analysis models (DAMs) would affect CDE standards and interoperability. caBIG® has used CDE standards as a mechanism to enable seamless interoperability (semantic and syntactic) of models on the caGrid® when users wished to aggregate data (e.g., via the dynamic structured query language (DSQL) on caGrid). Our small group had been charged with providing examples of how developers (inside and outside caBIG®) will achieve interoperability with the new CDEs of DAMs and ISO 21090 datatypes, and including metadata from both forms (CRFs) and services/applications (UML models). More recently, caBIG® has adopted the Service-Aware Interoperability Framework (SAIF)/Enterprise Conformance and Compliance Framework (ECCF). The Address Interoperability group will now use this use case to look at another question: Where would the Address Data Elements or concepts from standard terminologies fit in at the Computationally Independent Model (CIM; conceptual), Platform-Independent Model (PIM; logical) or Platform-Specific model (PSM; implementable) levels in the context of the CBIIT implementation of SAIF/ECCF? This concrete example can provide insight into how the interoperability can work across levels of specificity.

AUTHORS AND AFFILIATIONS: Salvatore Mungal, M.S.[1]; Mary Cooper, B.S.M.T.[2]; Riki Ohira ,Ph.D.[3]; Dianne Reeves, R.N., M.S.N.[4]; Mukesh Sharma , Ph.D.[5]; Lynne Wilkens, Dr.P.H.[6]; Brian Davis, Ph.D.[7], William Barry, Ph.D.[1] [1]Duke University; [2]SAIC; [3]BAH; [4]NCI CBIIT; [5]Washington University in St. Louis; [6]University of Hawaii; [7]3rd Millennium.

119. NET-based Clients + Services + Cloud for Cancer Bioinformatics

A key to continued success for the caBIG® will be interoperability. Since Sept 2008, as part of our caBIG® Architecture Workspace activities, we have designed and developed prototype clients and services for caGrid® based on .NET. The overall goal of this effort is not merely to provide a "second source" of existing technology/functionality, with identical functionality, but rather to leverage the unique capabilities contained in the wide range of existing and emerging technologies based on .NET. In this poster, we present a progress report describing many of the new capabilities that we have created in the past few months for .NET and caGrid, focusing on new model-based development based on .NET and Windows. First, we describe our open-source "plug-ins" for Visual Studio 2010. These plug-ins provide easy-to-use functionality tailored for the caGrid environment. Second, we describe new capabilities for caGrid to create services and/or keep data in Windows Azure, which is Microsoft's cloud platform. Third, we describe how we use the new SQL Server Modeling CTP to drive software development from model all the way to running clients and services (and the cloud). Altogether, we create new capabilities for caBIG® from device to laptop to server to cloud.

AUTHORS AND AFFILIATIONS: Marty Humphrey, Ph.D.[1] [1]Department of Computer Science, University of Virginia.

120. ISO 21090 Adoption: The Plan and the Tools

In an effort to streamline the design and development of software components that enable computable semantic interoperability across multiple technology platforms, Center for Biomedical Informatics and Information Technology (CBIIT) has adopted the Enterprise Conformance and Compliance Framework (ECCF). One requirement of this framework is that each attribute of a static information model should be bound to a robust data type specification that fully and unambiguously expresses the semantics of the attribute. CBIIT has selected ISO 21090 as its robust data type specification. The ISO 21090 data type

standard is based on the considerable effort over an extended period of years by a group of international healthcare informatics experts from the US, Canada, Australia, and Europe representing ISO, CEN, and HL7. The standard provides definitions for common concepts used in the healthcare domain. The definitions of the data types in the balloted ISO 21090 document allows for the data types to be modified to suit user needs, and HL7 provides rules for modifying the data types into a localization while claiming conformance to the standard. In an effort to ease the burden of interoperability, a single localization will be adopted for CBIIT-funded projects. The first set of CBIIT localizations has been defined for a majority of the data types. These localizations have been implemented into caBIT® tools and have been used to produce a service on the caGrid. For project teams, this requires an understanding of the new data types, the versioning policy, and the steps to be taken when the localization does not meet the needs of the project.

AUTHORS AND AFFILIATIONS: Libby Prince[1]; John Eisenschmidt[1]; Satish Patel[2]; Santhosh Garmilla[2]; Dan Dumitru[2]; Vijay Parmar[2]; Prasad Konka[3]; Ye Wu[3]; Xiaoling Chen[3]; Aynur Abdurazik[3]; Eugene Wang[3]; David Ervin[4]; Justin Parmer[4]; Scott Oster[5]; Shannon Hastings[5]; Steve Langella[5]; Ann Wiley[6]; Carolyn Klinger[6]; Bronwyn Gagne[6]; Sreenath Nampally[6]; Sichen Liu[7]; Avinash Shanbhag[8]. [1]Sapient; [2]Lantern Three; [3]Ekagra Software Technologies; [4]SAIC; [5]The Ohio State University; [6]The Ohio State University/Inventrio; [7]SAIC-Frederick; [8]NCI CBIIT.

121. Plugin Architecture for caBIG® Applications: Empowering the Adopter Community

Two of the major initiatives at caBIG® this year are the construction of applications as assemblies of cooperating services, and bringing outside developers more directly into the building of caBIG® applications. In our poster, we present an approach that addresses both themes: modularizing applications via an OSGi based plugin framework. This brings the ethos of a service-oriented approach, with the rigor of well defined interfaces, from being applied solely to inter-service integration, and into the heart of application design. At the same time, it simplifies code contribution by developers outside the core team, by eliminating hidden dependencies among application components and allowing contributions of individual components. An additional important benefit of this approach is the ability to easily combine functional modules to provide applications that present a more integrated user experience. Based on caArray, we present a prototype user interface that allows a researcher to locate data of interest and see summary data and visualizations for it in an integrated fashion. The visualizations are implemented as plugin widgets, based on the concept of "bringing analysis to the data." We describe the architectural challenges and lessons learned of our approach, as well as a blueprint for other applications interested in adopting a plugin architecture.

AUTHORS AND AFFILIATIONS: Dan Kokotov[1]; Rashmi Srinivasa[1]. [1]5AM Solutions, Inc.

122. Life Sciences Subject Matter Experts Model Business Processes and Capabilities to Support ECCF

The caBIG® Life Sciences Business Architecture Model 1.1 (LS BAM 1.1) is a foundational tool supporting caBIG®'s goal of providing an informatics framework for "bench to bedside and back" research. BAMs supply a common set of terminology, use cases, and actors for a domain, thus providing clear, concise descriptions of business capabilities, governance structures, business processes, and business knowledge. The LS BAM is a common reference point to communicate how Life Sciences (LS) research is performed. It can be used for tracing components of conceptual, logical, and implementation models to the standard business processes and capabilities and to inform high level requirements for

constructing software and services that use LS research data, model LS research practices, or provide services for LS researchers. Working iteratively with a Business Analyst, the LS Subject Matter Experts developed the LS BAM, describing goals that are common to all LS research sub-domains, including: how LS research is planned and performed; the organizations, resources, and methodologies involved; how data and materials are controlled and analyzed; and how outputs are disseminated. The primary output from the LS BAM effort has been a Use Case UML model (for business capabilities), with 90 use cases providing significant documentation about LS business goals through detailed storyboards, pre-conditions, the basic flow of events to accomplish the goal, and post-conditions. Activity Diagrams (for business processes) graphically illustrate the chronological or logical arrangement of these goals. Since Activity Diagrams tend to document very specific processes, they are created as areas of priority are identified, and a single exemplar was constructed. Lastly, actors are associated with use cases in Use Case and Activity Diagrams to convey ownership and responsibility. The LS BAM currently contains 61 actors, each representing a person, group, or institution.

AUTHORS AND AFFILIATIONS: Lauren Becnel Boyd, Ph.D.[1]; Scott P. Hunicke-Smith, Ph.D.[2]; Grace A. Stafford, Ph.D.[3]; Elaine T. Freund, Ph.D.[4]; Michele Ehlman[5]; Uma Chandran, Ph.D.[6]; Robert Dennis, Ph.D.[7]; Anna T. Fernandez, Ph.D.[8]; Stephen Goldstein[9]; David Steffen, Ph.D.[10]; Benjamin Tycko, M.D., Ph.D.[11]; Juli Klemm, Ph.D.[12] [1]Baylor College of Medicine; [2]University of Texas, Austin; [3]The Jackson Laboratory; [4]3rd Millennium; [5]Essex Management; [6]University of Pittsburgh; [7]University of California, Los Angeles; [8]Booz Allen Hamilton; [9]Sapient Government Services; [10]Baylor College of Medicine; [11]Columbia University; [12]NCI CBIIT.

123. Construction of a Standards-Based Platform-Independent Information Model for the Minnesota Congenital Heart Network

The Minnesota Congenital Heart Network (MCHN) is a group of multidisciplinary investigators dedicated to improving outcomes for children with congenital heart disease. To achieve these goals the federated systems in the MCHN must be interoperable. Therefore, a platform-independent model (PIM) was developed to facilitate clinical research data collection, analysis, and sharing. The PIM is a standards-based information model that utilizes resources from international standards organizations (such as NCI, HL7, CDISC, and the FDA), maximizing the potential of interoperability between the MCHN and other clinical data systems. The MCHN PIM was derived from the BRIDG domain analysis model (DAM) and was informed by the Mayo Clinic Enterprise Data Model. It is bound to the ISO 21090 health informatics datatypes (draft specification) and is expressed in the unified modeling language (UML). The MCHN PIM contains classes that represent individuals (patients), organizations, medical and surgical procedures, pharmaceuticals, diagnoses, and the results of laboratory tests and physical exams (vital signs). The MCHN PIM will allow members of the MCHN to exchange data effectively, and the standards-based process used to generate the model will facilitate interoperability with other clinical systems, including electronic health records and local, state, or national health information networks (e.g., the National Health Information Network, NHIN).

AUTHORS AND AFFILIATIONS: Robert R. Freimuth, Ph.D.[1]; Sheri Crow, M.D.[2] [1]Division of Biomedical Statistics and Informatics, Mayo Clinic; [2]Division of Pediatric Critical Care Medicine; Mayo Clinic.

124. Web-based Semantically-integrated Framework for Cancer Research

The UNMC Eppley Cancer Center has developed several multi-center web-based cancer-specific registries: Pancreatic, Breast, and Thyroid Cancer Collaborative Registries (PCCR, BCCR, and TCCR). These registries collect a variety of detailed data on cancer patients and individuals at high risk. Collected data includes demographic, lifestyle, quality of life, dietary preferences, physical activities, genetic, family and medical history, diagnostic procedures, treatment, etc. The user interface is compatible with all major web browsers and optimized for mobile devices, such as Apple iPads and iPhones. The PCCR has obtained a license for the caBIG® Bronze Compatible mark. Work on acquiring such license for the BCCR and TCCR is underway. To manage biospecimen data, the PCCR, BCCR, and TCCR have been coupled with the caTissue Suite. As of June 2010, data on more than 4,000 subjects have been collected in the registries and over 9,000 specimens have been managed by the caTissue. This data will serve as the foundation for a novel informatics framework that is currently under development. The framework will consist of two major components: the semantically integrated data repository (SIDR) and the data reporting and mining system (DRAMS). The SIDR will integrate and semantically harmonize de-identified data from the PCCR, BCCR, and TCCR with data from the Surveillance Epidemiology and End Results (SEER) database. The DRAMS will incorporate novel statistical models that we have been developing and will serve as an interface to the SIDR to allow querying data stored in this repository. The use of this framework will allow researchers not only to combine surveillance data available from the PCCR, BCCR, TCCR and SEER, but also increase statistical power for studies utilizing the collected data.

AUTHORS AND AFFILIATIONS: Oleg Shats, M.S.[1]; Alexander Sherman, M.S.[1]; Jianmin Feng, M.S.[1]; Tengiz Mdzinarishvili, Ph.D.[1]; Michael Gleason[1]; Leo Kinarsky, Ph.D.[1]; Simon Sherman, Ph.D.[1] [1]Eppley Cancer Center, University of Nebraska Medical Center.

125. Leveraging cancer Bioinformatics Infrastructure Objects (caBIO) for Research and Experimental Annotations

Conducting biomedical research requires access to experimental data as well as associated molecular annotations. Annotations providing detailed information on the molecular origin, biological process, and genetic alterations of associated experiment data can provide important insight on experimental outcomes. Annotating experiments with biological information requires access to data from a variety of disparate data sources in an integrative view. Cancer Bioinformatics Infrastructure Objects (caBIO) provide Application Programming Interface (API) access to an integrative view of molecular annotations originating from a variety of data sources. Information from key annotation providers including “UniGene, Entrez Gene, and UniProt,” is stored in the caBIO infrastructure and is updated on a semi-monthly basis. Recent additions to caBIO have included annotations from popular microarray platforms, curated information from the Cancer Gene Index (CGI), and pathway interactions from the Pathway Interaction Database (PID). caBIO data is made accessible through a variety of APIs including the caBIG® grid (caGrid) API and the RESTful, SOAP, and Java APIs. caBIO is also searchable through a variety of graphical user interfaces including: a caBIO Home Page which provides a utility (FreestyleLM) for performing Google®-like searches; a caBIO Portlet for easily accessing caBIO data via the caGrid® Portal by performing pre-defined templated searches; and a caBIO iPhone App for retrieving molecular annotations via the iPhone, iPod, and iPad devices. caBIG® translational research, knowledge

discovery, and genomic analysis tools like caIntegrator, Rembrandt, geWorkbench, and other biomedical applications leverage caBIO for genomic annotations. Use of the caBIO API avoids duplication of annotation data and reduces the need for additional data management resources supporting annotations. The caBIO project is currently engaged in a pilot effort involving the development of a molecular annotation service leveraging the NCI Enterprise Conformance and Compliance Framework (ECCF).

AUTHORS AND AFFILIATIONS: Juli Klemm, Ph.D.[1]; Avinash Shanbhag, Ph.D.[1]; Jim Sun[2]; Konrad Rokicki[2]; Sharon Gaheen, M.B.A.[2]; Liqun Qi[2]. [1]NCI CBIIT; [2]SAIC.

126. Development of an Enterprise Level Molecular Annotation Service Leveraging the NCI ECCF

caBIO (cancer Bioinformatics Infrastructure Objects) is a robust resource for accessing biomedical annotations from curated data sources in an integrated view in support of knowledge discovery. Currently, caBIO exposes data through APIs and interfaces including the Java, SOAP, RESTful, and Grid Data Services APIs. To provide enterprise-level services that facilitate semantic interoperability across the biomedical enterprise, the caBIO project is engaged in a pilot involving the development of a molecular annotation service that leverages the NCI's Enterprise Conformance and Compliance Framework (ECCF). The goals of the pilot are to: leverage caBIO as a reference implementation of the ECCF in support of the development of services that can achieve interoperability in a variety of deployment contexts; provide guidelines to assist other NCI CBIIT products in leveraging ECCF processes and developing ECCF artifacts; and identify tools and infrastructure that will assist in the development of services and specifications. Working towards these goals, the caBIO team has developed specifications and a project analysis model for a molecular annotation service. The molecular annotation service provides a standardized set of interfaces for querying information on the biological functions and classifications of a gene, genetic variants associated with a gene, the physical location of a gene, diseases and agents associated with a gene, and other annotations of interest. The molecular annotation service derives from the NCI enterprise domain models including the Life Sciences Domain Analysis Model (LS DAM) and the NCI's Biomedical Research Integrated Domain Group (BRIDG). Deriving from these models will assist in achieving enterprise level semantic interoperability and provide service consumers with a reusable set of interfaces for molecular annotations. Service interfaces were developed leveraging NCI CBIIT Service Development Tools supporting the NCI CBIIT localization of ISO 21090 data types. Information on the pilot effort is available on the caBIO Wiki: https://wiki.nci.nih.gov/display/caBIO/caBIO+ECCF.

AUTHORS AND AFFILIATIONS: Jim Sun[1]; Konrad Rokicki[1]; Sharon Gaheen, M.B.A.[1]; Liqun Qi[1]; Baris Suzek[2]; Brian Davis, Ph.D.[3]; Elaine Freund, Ph.D.[3]; Juli Klemm, Ph.D.[4]; Avinash Shanbhag, Ph.D.[4]; [1]SAIC; [2]Georgetown University; [3]3rd Millennium; [4]NCI CBIIT.

127. Using GEM to Foster Sharing and Collaboration on the Grid

Over the past few years, the National Cancer Institute (NCI) in collaboration with Westat, has developed the Grid-Enabled Measures (GEM) database. GEM was conceived to use collaborative web technology to facilitate two primary goals for behavioral researchers: (1) Promote the use of standardized measures which are tied to theoretically-based constructs; and (2) Facilitate the ability to share harmonized data resulting from the use of standardized measures. While these activities are currently being routinely conducted in the clinical areas, these efforts represent early

attempts to leverage the resources of caBIG® and the caGrid® platform for behavioral researchers. The purpose of this poster is to describe how GEM was used as a prototype for developing standardized practices and procedures for promoting sharing of data, metadata, and measures in the behavioral science research area. The steps necessary to move beyond a conceptual framework for GEM development included four major activities. First, informational interviews were conducted with researchers in the tobacco cessation area to assess what types of data (e.g., measures, constructs) and metadata (e.g., reliability, validity) would be valuable to share among researchers in their field. Second, based upon their feedback, the GEM team developed a structure for displaying the information in a user friendly fashion. Third, a prototype was developed and tested using a group of researchers identified by NCI. Fourth, based upon the feedback received, a beta version of the site was developed to include the opportunity for researchers to use GEM to create consensus to promote the use of standardized measures, their theoretically-based constructs and publically available datasets on the grid. While the GEM site continues to be improved and expanded, the goal remains the same--to promote collaboration and research efficiency.

AUTHORS AND AFFILIATIONS: Kerry Levin, Ph.D.[1]; Cynthia Helba, Ph.D.[1]; David Garner[1]; Marsha Dunn[1]; Kisha Coa[1]; Keith Umbel[1]; Richard Moser, Ph.D.[2]
[1]Westat; [2]National Cancer Institute.

128. caLIMS2: Next Generation Cancer Laboratory Information Management System Designed to be caBIG® Interoperable

The purpose of the caLIMS2 project is to create a Laboratory Information Management System (LIMS) that is interoperable within established caBIG® standards and guidelines. Currently, caBIG® tools exist for Clinical Trials implementation and monitoring, for biospecimen storage and tracking, for data storage and retrieval, and for data analysis. However, a gap exists in the caBIG® tool set related to the capture and tracking of laboratory activities. caLIMS2 will fill that gap - tracking a complete laboratory workflow that uses materials from a specimen management service (e.g. caTissue) to generate experimental results for one of the caBIG® data management services (e.g. caArray). Core caLIMS2 functions are organized in four basic modules (administration, inventory, workflow, and reports) and include the management of personnel, equipment, lab supplies and reagents, samples, laboratory workflow, and experimentally derived metadata and data. caLIMS2 is highly flexible, making it suitable for research labs and high throughput core facilities, and for support of multiple research domains (genomics, proteomics, nanoparticle characterization, etc.). caLIMS2 will help further translational cancer research through the organization of laboratory workflow, tracking of specimens and derived specimens, acquisition of laboratory data and metadata associated with those specimens, and the appropriate sharing and dissemination of the data to support subsequent analyses. caLIMS2 v1.0 focuses on basic functionality and easy integration of caBIG® specimen management and data management services. We will present the caLIMS2 v1.0 UML model, a prerelease package, and describe our progress in developing the version 1.0 release. The caLIMS2 v1.0 design includes integration with other caBIG® application services (e.g. caArray and caTissue) and the creation of two enterprise level services: Equipment Service and Experiment Method Service. More details and information on the project can be found on the caLIMS2 Project Wiki site (https://wiki.nci.nih.gov/x/2oMYAQ) and the caLIMS2 GForge site (http://gforge.nci.nih.gov/projects/calims2/).

AUTHORS AND AFFILIATIONS: Jenny Kelley, M.A.[1]; Robert Clifford, Ph.D.[1]; Sasikumar Thangaraj, M.S.[2]; Cu Nguyen, B.S.[1]; Henry Zhang,Ph.D.[1] Anand Basu, M.S.[3]; Juli Klemm, Ph.D.[3]; Kenneth Buetow,

Ph.D.[3] [1]NCI Laboratory of Population Genetics; [2]Moxie Informatics; [3]NCI CBIIT.

129. The TCR Cancer Registry: A Case Study of caCore Based Data Standards Implementation to Integrate with the Cancer Biomedical Informatics Grid®

Background: caBIG® is a network of information and tools. It allows institutions and individuals to easily share, integrate, and analyze data. A fully interoperable IT network has been developing by caBIG® that will supercharge the entire field of cancer research. The Taiwan Cancer Registry (TCR), established in 1979, is organized and funded by the Health Department of the central government. Hospitals with a capacity of more than 50-bed capacity are recruited to participate in reporting all newly diagnosed malignant neoplasms to the registry. The TCR is assisted by a Cancer Registry Advisory Board, comprising 16 expert members from various fields. The TCR's primary goal is to survey the incidence of cancer in Taiwan. The NCICB has created a core infrastructure called caCore. We developed a syntactically and semantically interoperable biomedical information system, which confirmed the caCore framework. The framework has three key components: (1) Enterprise Vocabulary Services (EVS), (2) cancer Data Standards Repository (caDSR), (3) caCore SDK and associated tools for model-driven software engineering of systems. Because currently there is no electronic data standard for TCR, developing a nationwide standard format and confirming to caBIG® is essential to cancer research society in Taiwan. Methods: In this study, we provide a case study of caCore based data standard implementation of the TCR, which is a population-based database. By following the building process for caCore-based services, we expect to create our TCR node on the Grid in six steps: (1) creating UML model of TCR; (2) semantic annotation; (3) transforming UML model into metadata; (4) SDK Code Generation; (5) Compatibility Reviewing; and (6) creating a Node on the Grid. Result: A prototype of caCore-based TCR data standard has been established. The TCR cancer registry is a basis of electronic data standards in cancer research using caBIG® semantic modeling methodology.

AUTHORS AND AFFILIATIONS: Shin-Bo Chen[1]; Jeng-Fong Chiou[2]; Wen-Ta Chiu[3]; Hung-Wen Chiu[4]; Yuan-Chii Lee[4]; Chien-Yeh Hsu[4]. [1]Graduate Institute of Medical Informatics, Taipei Medical University; [2]Cancer Center and Section of Hematology-Oncology, Department of Internal Medicine, Taipei Medical University and Hospital; [3]Department of Neurosurgery, Graduate Institute of Injury Prevention and control and Center of Excellence for Cancer Research (CECR), Taipei Medical University; [4]Graduate Institute of Medical Informatics and Center of Excellence for Cancer Research (CECR), Taipei Medical University.

130. caEHR Information Management - A Framework for Artifact Development and Deployment

The Information Management team will implement a dynamic artifact development and deployment system to provide robust content integrity as well as a streamlined delivery mechanism of the caEHR artifacts. The development cycle will include the implementation of a Content Component Management System (CCMS) to support Extensible Markup Language (XML), Topic-based authored information that will exploit the retrieve, reuse, and re-purpose abilities of the information that is developed by the caEHR Analyst, Q/A, and Architect teams. Our primary feature will include the implementation of the Darwin Information Typing Architecture (DITA), Desktop Type Definition (DTD).The delivery mechanism will be supported by an Enterprise Content Management (ECM) system to serve as a metadata-driven repository for internal caEHR and other potential end users. The five primary ECM components and technologies will include: Artifact capture, management, storage, preservation, and delivery.

AUTHORS AND AFFILIATIONS: Brian Schmitt[1]; Steve Manning[1]. [1]Ekagra Software Technologies.

131. Setting a New caBIG® "Openness Standard" - Building a Better "Network"

The session illustrates transformation of the information gathering process and collaborative nature of communities of practice within the infrastructure of caBIG®. This infrastructure will provide access in a Single-Sign-On environment to caBIG® 's Community Website, the Knowledge Center Web Infrastructure, the NCI Confluence Wiki, Gforge, the LMS and Training Server as well as the Public Site. The ease to find, manipulate, and disseminate information and data, will allow researchers and policy makers alike to revolutionize respective processes, improving quality and productivity, while maintaining a high level of collaboration and discovery. MB facilitates the creation of communities of users and practice to interact, share best practices, and learn from each other outside of formal classes and sessions across the multiple repositories of caBIG®. AcrossWorld will illustrate cases of its global approach to share and share-a-like content. The project facilitates development of a research and production culture, encouraging dialogue and co-creation outside traditional networks, while exposing access to information geographically and repository dispersed. Researchers, doctors, and nurses, can focus on new development and methodology. People are more readily available, using time more efficiently, as the tools and processes will allow for innovation and collaboration, as documents, materials, and resources can be easily worked on in formal and informal network across networks. These tools and processes enable the ready access of information and data in ways not previously utilized within caBIG®. The holistic integration of e-portfolio level details and the contextualization of content and user relevant data combined with the option for multi-modal delivery methods provide caBIG® a more systemic and programmatic approach by using channel delivery options best suited to specific audiences

AUTHORS AND AFFILIATIONS: Stephan K. Thieringer, Ph.D.[1]; Peter Lamothe[1]; Larry Lessin[2]; Gerry Hanley[3]. [1]AcrossWorld Education; [2]Open Educational Resources Cancer; [3]California State University.

132. caGrid WebSSO

WebSSO is an enterprise Single Sign-On/Sign-Out solution for caGrid based web applications. Single Sign-On (SSO) provides a better user experience when running a multitude of multiple web applications with their own means of authentication. In a WebSSO enabled solution, the authentication process is delegated to a central authoritative source of trust. Once the user logs in to the WebSSO server, they are not required to log in to separate caGrid applications. It provides an automated mechanism for delegation and retrieval of the user's caGrid credentials thereby avoiding the transfer of the caGrid credentials among caGrid applications. WebSSO is built on top of the JA-SIG central Central authentication Authentication service Service (CAS) and caGrid Authentication and Authorization with Reliably Distributed Services (GAARDS) framework to provide the core Single Sign-On/Sign-Out capabilities.

AUTHORS AND AFFILIATIONS: Santhosh Garmilla[1] Kunal Modi[1]; Steve Langella[2]; Scott Oster [2]. [1]Ekagra Software Technologies; [2]Inventrio.

133. Common Security Module v4.2

CSM provides a flexible, comprehensive solution to common security objectives. Development teams can use CSM services rather than creating their own independent security methodology. CSM uses Java Authentication and Authorization Standard (JAAS) to authenticate the users and integrates easily with other frameworks or

standards. The features of CSM include authentication, authorization, authorization policy provisioning, and role/group/object/instance/attribute-level security. It consists of the APIs for programmatic integration, User Provisioning Tool (UPT) for web-based authorization (security) policy management, and CSM GAARDS Migration Module (CGMM) as an intermediate interface to migrate local users to caGrid.

AUTHORS AND AFFILIATIONS: Vijay Parmar[1]; Sichen Liu[2]; Avinash Shanbhag[2]. [1]Ekagra Software Technologies; [2]NCI CBIIT.

134. CSM GAARDS Migration Module (CGMM) v0.6

The CGMM was chartered to provide a solution to migrate existing (web or services) applications from Common Security Module (CSM)-based authentication to Grid Authentication and Authorization with Reliably Distributed Services (GAARDS)-based authentication. CGMM allows applications to avoid duplication of accounts, provides single set of credentials that can be used for multiple applications and provides user-friendly provisioning of new Users with Grid Identities. CGMM comes with configurable caGrid Identity providers for authentication and also leverages configured caGrid Authentication services and Dorian services. CGMM Web also provides high configurability in the form of default vs. alternate behavior, ability to operate in Standalone mode, and other miscellaneous workflow configurations.

AUTHORS AND AFFILIATIONS: Vijay Parmar[1]; Sichen Liu[2]; Avinash Shanbhag[2]. [1]Ekagra Software Technologies; [2]NCI CBIIT.

135. Instance and Attribute Level Security: High Performance, Out-of-Box Solution From CSM

Common Security Module (CSM) v4.2 provides high performance, data tier-level instance and attribute security. Instance-level security provides direct and cross-dependent security where access to an instance of object is controlled using the instance itself or an associated object. Attribute-level security provides access controls to the object attributes which can be managed based on user access privileges. The instance and attribute level security feature is available out of box with caCore SDK generated systems and is easy to adapt for non-SDK systems.

AUTHORS AND AFFILIATIONS: Vijay Parmar[1]; Satish Patel[1]; Sichen Liu[2]; Avinash Shanbhag[2]. [1]Ekagra Software Technologies; [2]NCI CBIIT.

136. Service Development Tools for ISO 21090 Datatypes

In its continued pursuit of easing the burden for publishing cancer-related information, CBIIT has updated its service development tooling to support the first version of the CBIIT-defined localizations of the ISO 21090 Standard. This effort involved three development teams whose efforts extended caAdapter, caCore SDK, and caGrid's Introduce toolkit. This product extension enabled development of applications and grid services that support complex ISO 21090 datatypes within a Model-Driven Architecture (MDA). This poster will provide an update of the CBIIT policy, an overview of the technical solution, and a description of the impact on the tooling from a user's perspective.

AUTHORS AND AFFILIATIONS: Satish Patel[1]; Santhosh Garmilla[1]; Dan Dumitru[1]; Vijay Parmar[1]; Prasad Konka[2]; Ye Wu[2]; Xiaoling Chen[2]; Aynur Abdurazik[2]; Eugene Wang[2]; Libby Prince[3]; John Eisenschmidt[4]; David Ervin[5]; Justin Parmer[5]; ,Scott Oster[6]; Shannon Hastings[6]; Steve Langella[6]; Ann Wiley[7]; Carolyn Klinger[7]; Bronwyn Gagne[7]; Sreenath Nampally[7]; Sichen Liu[8]; Avinash Shanbhag[8]. [1]Ekagra Software Technologies; [2]SAIC; [3]Sapient; [4]Lantern Three; [5]The Ohio State University; [6]TheOhio State University/Inventrio; [7]SAIC-Frederick; [8]NCI CBIIT.

137. Single Sign-On/Sign-Out for caGrid Enabled Web Applications

WebSSO is an enterprise Single Sign-On/Sign-Out solution for caGrid based web applications. Single Sign-On (SSO) provides a better user experience when running multiple web applications with their own means of authentication. In a WebSSO enabled solution, the authentication process is delegated to a central authoritative source of trust. Once the user logs into the WebSSO server, they are not required to log in to separate caGrid® applications. It provides an automated mechanism for delegation and retrieval of the user's caGrid® credentials thereby avoiding the transfer of the caGrid® credentials among caGrid® applications. WebSSO is built on top of the JA-SIG Central Authentication Service (CAS) and Grid Authentication and Authorization with Reliably Distributed Services (GAARDS) framework to provide the core Single Sign-On/Sign-Out capabilities.

AUTHORS AND AFFILIATIONS: Santhosh Garmilla[1]; Kunal Modi[1]; Steve Langella[2]; Scott Oster[2]; Joshua Phillips[3]. [1]Ekagra Software Technologies; [2]The Ohio State University; [3]Semantic Bits,LLC.

138. caCore Workbench

The caCore Workbench is a part of NCI CBIIT's integrated service development suite supporting NCI CBIIT Localization of ISO 21090 datatypes. It facilitates management of code generation and deployment configuration for a project using a simple GUI. Using this workbench, the user can create a new data service application by providing the UML model and code generation settings. The user can use the deployment workflow of the Workbench to deploy the generated system on a local or remote machine with only a few clicks. In addition, the Workbench provides users with step-by-step validation of configuration parameters. The Workbench saves time and reduces human error in generating a semantically integrated caCore-like system, enabled with the ISO 21090 datatypes.

AUTHORS AND AFFILIATIONS: Satish Patel[1]; Santhosh Garmilla[2]; Dan Dumitru[1]; Prasad Konka[2]; Ye Wu[2]; Xiaoling Chen[2];Sichen Liu[3]; Avinash Shanbhag[3]. [1]Ekagra Software Technologies; [2]SAIC; [3]NCI CBIIT.

139. caCore Software Development Kit (SDK)

caCore Software Development Kit (caCore SDK) is a part of NCI CBIIT's integrated service development suite supporting NCI CBIIT Localization of ISO 21090 datatypes. It generates a data management framework to allow researchers to navigate through a large number of data sources. A caCore SDK generated middleware system is built using the principles of Model Driven Architecture (MDA), n-tier architecture, and common APIs. This middleware system allows for easy access to data by other applications in either a secured or non secured fashion. When the caCore SDK generated system is combined with controlled vocabularies and registered metadata, the resulting caCore-like system, enabled with ISO 21090 datatypes, is then considered semantically integrated. All exposed API elements of the generated system have runtime-accessible metadata that defines the elements using controlled terminology.

AUTHORS AND AFFILIATIONS: Satish Patel[1]; Santhosh Garmilla[2]; Dan Dumitru[1]; Vijay Parmar[1]; Prasad Konka[2]; Ye Wu[2]; Xiaoling Chen[2];Sichen Liu[3]; Avinash Shanbhag[3]. [1]Ekagra Software Technologies; [2]SAIC; [3]NCI CBIIT.

140. Accessing and Consuming caBIG® Training

The cancer Biomedical Informatics Grid® (caBIG®) offers more than 120 open source tools, technologies and infrastructure components that address the needs of the biomedical research community to share data and knowledge, simplify collaboration,

and speed research. Did you know that many of these have training and other educational materials? Delve into the caBIG® Tool Pages, Knowledge Centers, Wikis, and Learning Management System (LMS) to access the wealth of information that caBIG® has to offer. Enroll in courses and track your learning progress through the LMS. Access demonstrations and documentation, and interact with members of the community on the Knowledge Centers. Find the tools you need and learn about their capabilities on the caBIG® Tool Pages. Contribute to the caBIG® knowledge base and discover new learning opportunities by accessing the Wikis. Training and other educational content is organized by subject area, tool, and role, making it easier for learners to find materials relevant to their background and needs. Varied learning opportunities and media put community members in charge of their own learning by giving them control over the 'what, who, how, why, when, and where' of their educational experiences in caBIG® creating convenient and realistic options to access and use the rich collection of educational content caBIG® offers.

AUTHORS AND AFFILIATIONS: Jennifer Brush[1]; Dianne Reeves, R.N., M.S.N.[2]; Leslie Derr, Ph.D.[2]; Jamie Parker[1]; Jennifer Tucker, Ph.D.[3]; Calla Pearce[4]; Lori Thompson, R.N., B.S.N.[1] [1]ScenPro, Inc.; [2]NCI CBIIT; [3]Otto Kroeger Associates; [4]Terrapin Systems, LLC.

141. caBIG® Learning Management System: A Tool for Getting Connected with caBIG®

This poster will show how the Learning Management System can help users get connected with caBIG® by identifying training classes and role-based programs designed to help them learn about caBIG® and the tools available to them. The poster will focus on how everyone from caBIG® Newcomers to seasoned participants can find information about all aspects of caBIG® through the LMS. It will show how users can access training materials and how they can create and LMS account if they need to track required classes.

AUTHORS AND AFFILIATIONS: Calla Pearce, M.A.[1] [1]Terrapin Systems, LLC.

142. caBIG® caGrid® Knowledge Center

The caGrid Knowledge Center, part of the caBIG® Enterprise Support Network, is an expert resource on Grid technologies available to the broader biomedical informatics community. Our primary mission is to provide tools and resources to enable collaborative e-Science. We provide caGrid software and documentation to the broader Grid community, collaborate with scientific projects to utilize Grid technologies, maintain a Community Training Grid for testing purposes, and support a Community Projects effort to facilitate caGrid-compatible software re-use. In addition to caGrid platform expertise, we facilitate broad-based adoption and adaptation of caBIG®'s SOA components to address driving biological or clinical problems. Our poster provides a brief introduction of the caGrid Knowledge Center and the Enterprise Support Network. We present a listing of the Knowledge Center's institutions and personnel, an overview of the Community Training Grid, and highlight our Community Projects effort, which is focused on fostering emerging Grid projects, and providing reusable, open-source Grid technologies to the broader caGrid community. The focus of our poster is on caBIG® adopters and adapters, describing the key aspects of their deployments and highlighting relevant lessons learned. Example adopters and adapters are communities (e.g., CTSA) as well as institutions and individuals, such as University of Minnesota and others. Reference: Scott Oster, Stephen Langella, Shannon L. Hastings, David W. Ervin, Ravi Madduri, Joshua Phillips, Tahsin M. Kurc, Frank Siebenlist, Peter A. Covitz, Krishnakant Shanbhag, Ian Foster, Joel H. Saltz, "caGrid 1.0: An Enterprise Grid Infrastructure for Biomedical

Research",Journal of the American Medical Informatics Association, Vol. 15(2): pp.138-149, December, 2007.

AUTHORS AND AFFILIATIONS: Justin Permar[1]; William Stephens[1]; Joe George[1]; Albert M. Lai, Ph.D.[1]
[1]Center for IT Innovations in Healthcare, Department of Biomedical Informatics, The Ohio State University.

143. Ushering in Change, the caBIG® Way

Over the past few decades, rapid developments in genomic and other molecular research technologies and developments in information technologies have combined to produce a tremendous amount of information related to molecular biology. This has also led to change in the way in which clinical and bio informatics research is performed. Research organizations have started leveraging the benefits of ever-growing IT infrastructure and applications for improvements in research and analysis. Initiatives like caBIG® have acted as a catalyst by providing a network of infrastructure, tools, and ideas that has further enabled collection, analysis and sharing of data and knowledge. This has generated a vast amount of data with greater opportunities in using patient data and medical knowledge to improve quality and efficiency of care.The proliferation of bioinformatics activities brings new challenges: how to understand and organize these resources, how to exchange and reuse successful experimental procedures, and to provide interoperability among data and tools. In some instances this shift has altered core ways of conducting business within organizations by introducing changes in processes and procedures to suit the current systems, which in turn has generated resistance within professionals for adopting new systems. Similarly while adopting new applications, future benefits are not seen sufficient to outweigh the short-term investment required to learn new programs. Thus our main challenge has been to implement caBIG® concepts and systems as smoothly as possible to enhance the quality of work life and take into consideration the organizational goals without comprising on the caBIG® vision. The objective of this poster is to demonstrate real examples of smart and innovative integrations that we have been able to offer to some of the organizations leveraging caBIG® suite of applications. The poster will also emphasize on semantic integration of caBIG® applications with other non-caBIG® tools and utilities.

AUTHORS AND AFFILIATIONS: Harshal Shah[1]; Paul Perapadan[1]. [1]Persistent Systems Limited.

144. Assembling Biomedical Collaborative Intelligence Using Google Wave Federation Protocol

Extensive Markup Language (XML) has ushered in a new openness to information technologies. With XML, various entities can now inter-communicate and inter-process disparate forms of data, allowing entities to freely exchange ideas that were previously represented in various formats. Since this universal format is malleable, software tools can be created to provide and promote collaboration, along with providing the ability to shape, process, and represent those data beyond text. Using XML as a foundation, Google has launched a mechanism (Wave Federations Protocol) that seeks to change the way the world traditionally uses electronic communications. Google Wave gives collaborators the ability to graphically create a Wave (a behind the scenes XML-based communication document) and invite others to join that Wave. A Blip gives the user multifaceted capability to include instant messaging, drag-and-drop documents, PDFs, DICOM images, non-identifiable HL7 files, audio, or video. All user interaction is real time, even on-the-fly translation to over 40 languages. These collaborative dynamics lend themselves to easily develop flexible workflows, especially in the area of biomedical research. For instance, caBIG® applications can integrate

with this technology to exceed the limits of disparate data domains and storage mechanisms. The rapid and flexible nature of XML gives caBIG® tools the capability to interoperate with the Wave Federation Protocol, instead of requiring the project to have a certain level of Java-savvy expertise. Successful integration of Shibboleth-generated SAML2 and caBIG® GAARDS provides an extremely secure collaboration environment. Additionally, the Google Wave contains a playback feature that allows newly-on boarded collaborators to obtain a quick, yet thorough project history to accurately reflect the prior decisions that shaped the project's vision. The Google Wave Federation Protocol, the GLOBUS toolkit and caBIG® are key partners in an effective collaborative biomedical pipeline.

AUTHORS AND AFFILIATIONS: William J. Girten, Sr.[1]; Luan Le[1]. [1]BioGadgets, LLC.

145. VIVO and eagle-i: Why They are Important for caBIG®

Being able to share data and resources in a networked environment that semantically interoperates can promote collaboration among cancer researchers. caBIG® provides a grid computing environment which allows cancer researchers to share data and analytical services. The grid infrastructure for caBIG®, caGrid, is based on semantic service oriented architecture (sSOA), which provides a means for datasets or analytic services to advertise their presence to be found by prospective users. In 2009, the NIH's National Center for Research Resources (NCRR) awarded two new ARRA grants, VIVO and eagle-i, which will provide a complementary way for cancer researchers to collaborate. VIVO is building a social network of researchers, their publications, and scientific activities, in a federated system for research networking, while eagle-i is creating a semantic repository and search engine for "invisible" resources in core facilities and research labs for resource discovery. These resources include equipment, animal models, software, protocols, and biological reagents such as cell lines and tissues. Both projects include a semantic infrastructure based on ontologies, which are being developed cooperatively by members of the two projects. VIVO and eagle-i will enable researchers to find each other and to discover research resources at participating institutions. When these researchers are at institutions that provide caBIG®, or BIG Health, they will also be able to take advantage of the caGrid infrastructure to share data and analytic services. NCRR will have a proof-of-concept system when the two projects are completed in August 2011. For the poster presentation, we will illustrate aspects of VIVO and eagle-i, how they are connected, and why they are important to caBIG®.

AUTHORS AND AFFILIATIONS: Leslie D. McIntosh, Ph.D.[1]; Paul Thompson, Ph.D.[2]; Mike Conlon, Ph.D.[3]; Lee Nadler, M.D.[4] [1]Washington University School of Medicine; [2]Dartmouth Medical College; [3]University of Florida; [4]Harvard Medical School; VIVO Collaboration; eagle-i Collaboration.

146. caBIG® Vocabulary Knowledge Center

The NCI caBIG® Vocabulary Knowledge Center (VKC) supports tools and infrastructure related to terminology storage, authoring, and editing. Specifically, the VKC provides a centralized resource for documentation and training materials, discussion forums, bug reports/feature requests, and source code for an integrated set of tooling that includes: (1) LexEVS, a terminology server that forms the foundation of the NCI Enterprise Vocabulary Services; (2) LexWiki, a framework for collaborative terminology authoring; and (3) NCI Protg, an ontology editor used for managing and implementing terminology change proposals. In addition to serving as the primary resource for information about the NCI terminology environment, the VKC also provides information about the broader NCI semantic infrastructure, as well as related

tooling (e.g., metadata harmonization, natural language processing), standard specifications (e.g., Common Terminology Services 2), standards organizations (e.g., HL7, OMG, ISO, W3C), and emerging technologies (e.g., semantic web). This information supports the development of systems that achieve working interoperability within the context of a federated and heterogeneous semantics environment.

AUTHORS AND AFFILIATIONS: Robert R. Freimuth, Ph.D.[1]; Troy Bleeker[1]; Christopher G. Chute, M.D., Dr.P.H.[1] [1]Division of Biomedical Statistics and Informatics, Mayo Clinic College of Medicine.

147. C-DAC: CBIIT Collaboration - Areas of Interest

A part of the collaborative effort between C-DAC and CBIIT, is cooperation on research, development, and deployment of biomedical informatics and healthcare-related information technology. The Medical Informatics Group of C-DAC has been active in the areas of ICT in healthcare, telemedicine, EHR/EMR, imaging, and cutting-edge computing and delivery/deployment infrastructure such as distributed systems and grid. In order to reduce introduction and maintenance cost of adoption of better-suited tools, it plans to utilize its expertise in customizing and enhancing existing caBIG® tools to grid infrastructure of C-DAC. C-DAC is also working in the area of Integrative Medical Informatics, which includes R&D for promotive, preventive, and curative healthcare. The areas of common interest include tools development using fundamental principles of Ayurveda (Indian System of Medicine) harnessing IT, medical instruments, and CAM. The spectrum covers sub-areas like the development of non-invasive medical instrument for early detection of Cancer from pulse morphology pattern/s, pre-clinical, clinical research, and drug discovery. Besides the above, the Applied Artificial Intelligence (AAI) group of C-DAC has been a pioneer in the area of developing technologies for machine translation in various Indian languages. This expertise will be utilized to deal with various terminologies used in cancer care in many Indian languages. This will effectively assist in taking caBIG® technologies to interface with the common man in many local languages. The Bioinformatics Resources and Applications Facility (BRAF) at C-DAC is an effort towards providing high-end supercomputing facility to the researchers working in area of Bioinformatics. The Bioinformatics group at C-DAC has expertise in the areas of Computational Genome Analysis and Advanced Molecular Simulations and also develops software towards solving problems in biology. The above expertise will be utilized to carry out advanced molecular dynamics simulations on cancer-related proteins. Results of such simulations would provide valuable insight to the researchers.

AUTHORS AND AFFILIATIONS: Hemant Darbari[1]; Pradeep Sinha[1]; Medha Dhurandhar[1]; Rajendra Joshi[1]; Ajai Kumar[1]; Gaur Sundar[1]. [1]Centre for Development of Advanced Computing (C-DAC), India.

148. User Interfaces: Designing Software to Look Like It REALLY Works Together

When we get new software, one of the first things we see is that it has a similar “look and feel” to other software from the same company. By the same token, experienced users can tell which company has desinged a particular package by it’s “look and feel” after they become familiar with that company’s User Interface standard. Helping caBIG® designers and those programmers involved in developing both designs and reference implementations to have these same properties is the objective here. This poster will teach readers how to think about User Interfaces, both in their historical context and contemporary usage, including: (1) What the international (ISO) and industry standards are today with examples; (2) Where the standards seem to be going with explanations; (3) Why caBIG® needs to build a User Interface standard with

examples; (4) How caBIG® can achieve a User Interface standard for its projects; and (5) What individuals and institutions can do to participate. With recent editorials in professional journals about the difficulty even skilled individuals have in negotiating forms and other online data entry mechanisms, it seems like the time is right for caBIG® to step up and implement a common User Interface standard for its activities.

AUTHORS AND AFFILIATIONS: Virginia R Hetrick, Ph.D.[1]; members of the User Interface BoF Sessions[1]. [1]caBIG® Patient Advocate

4b. Exhibitors

World's Fair Exhibitors – Hall C

(To learn more about the caBIG® Support Service Provider Program please refer to Tab 4c: ESN.)

5AM Solutions

A licensed caBIG® Support Service Provider

5AM Solutions' goals align with those of the caBIG® initiative. We look forward to and work toward the day when access to complex information is simple, when scientists have tools to help connect the dots, and when researchers, doctors, and patients tackle disease together. We believe that's achievable—and we're guaranteed to see more success from striving than from only hoping. That's why we're pleased to participate as a caBIG® Support Service Provider—we're ready to help organizations adapt, enhance, and create caBIG®-compatible software. We'll help you connect your existing software to caGrid. We'll pragmatically adapt and modify caBIG® tools to meet your institution's needs. We're professional software engineers who know the domain, the data, the technology, and the caBIG® initiative. Engage us as consultants, developers, or as part of your team.

CONTACT NAME: Leslie Power
E-MAIL: lpower@5amsolutions.com
WEBSITE: www.5amsolutions.com

Brady Corporation

Brady Corporation is a global manufacturer and supplier of labeling solutions for biological specimens with a focus on preprocess labeling of tissue cassettes, cryo cassettes, cryo vials, and slides. Brady labels are specifically designed for applications in the area of pathology and will support sample traceability from surgical bedside to long-term storage. Brady printers, labels, and scanners can and have been integrated with software applications and tools that have been certified as caBIG® Bronze compatible.

CONTACT NAME: Julie Schoenborn
E-MAIL: Julie_Schoenborn@bradycorp.com
WEBSITE: www.bradyid.com/specimenlabeling

CBIIT Training Team

As members of the CBIIT caBIG® training team, we provide a wealth of educational materials, training, and support that can be found on the caBIG® Community Website, caBIG® Tool Pages, Knowledge Centers, Wikis, and through the Learning Management System (LMS). Our goal is to support deployment, consumption, adaption, and adoption of caBIG® tools, infrastructure, technologies, and program objectives. We offer educational materials, including training, for many of the 120 open source tools, supported technologies, underlying infrastructure components, and program objectives that address the needs of the biomedical research community.

CONTACT NAME: Jenny Brush
E-MAIL: brushj@mail.nih.gov
WEBSITE: wiki.nci.nih.gov/x/OAFy

CTIS, Inc.

A licensed caBIG® Support Service Provider

CTIS is a health informatics company focused in the areas of Clinical Trial Research and Management (CTRM), Health Information Exchange (HIE), and personalized medicine solutions. CTIS is committed to making a difference in the health industry by converging clinical research, clinical trials, and translational research to develop a new paradigm of highly personalized and comprehensive chronic diseases such as cancer, heart, diabetes, neurology, and kidney disease;

information-based healthcare solutions that deliver tailored and targeted treatments for patients and a better outcome of care process. CTIS's primary focus areas include individualized healthcare; and health disparities. Our solutions help deliver improved quality of care, safety, and timeliness for our clients. CTIS offers end-to-end, enterprise-wide infrastructure, clinical applications, data warehousing, and business intelligence solutions to provide the right information to the right stakeholders, in the right place, at the right time. CTIS also offers mobile platforms that provide real-time access to patient outcome data at the point-of-care, allowing providers to make informed decisions efficiently and accurately, resulting in improved patient care and treatment outcomes. Our core solution includes clinical trial research management, health information exchange, chronic diseases and personalized medicine, and global health solutions. Our application offerings include data warehousing and data mining, process re-engineering and systems integration, and prevention and epidemiology. CTIS brings over 10 years of experience and expertise of building mobile applications while working with multiple personal health record vendors and research organizations. For over 20 years, CTIS has designed and implemented health information technology systems that incorporate seamless collaboration within a sustainable and scalable model to integrate and exchange data from a variety of sources. As a proven thought leader, CTIS utilizes business process re-engineering and organizational alignment to increase process efficiency, organizational effectiveness, and stakeholder productivity that contribute to the highest return on investment (ROI) and patient health outcomes.

CONTACT NAME: Crystal Evans
E-MAIL: cevans@ctisinc.com
WEBSITE: www.ctisinc.com

Daedalus Software, Inc.

Daedalus Software, Inc., is a Cambridge, MA-based biobanking automation company. Our product portfolio, Biomaterial Tracking and Management (BTM) application has a fully mapped data model with caTissue. The interface between BTM and caTissue allows exporting data from the BTM system into caTissue and to make it available on the grid. This has allowed cancer centers to send clinical data across the Grid for national studies such as for the SPORE projects.

CONTACT NAME: Azita Sharif
E-MAIL: asharif@daedalussoftware.com
WEBSITE: www.BTM-BINDS.com

Dataworks Development, Inc.

Dataworks Development is a software development firm providing database software, sample management, and bar code labeling solutions for biorepositories, pharmaceuticals, and health science research laboratories worldwide. Its flagship program, Freezerworks Unlimited, is a fully configurable sample management program in use by a number of cancer research organizations and has been certified as caBIG® Bronze compatible. All software is written according to FDA Guidelines for software development.

CONTACT NAME: Rick Michels
E-MAIL: rick@dwdev.com
WEBSITE: www.freezerworks.com

Deep Archive Checkpoint Technologies

Deep Archive Checkpoint Technologies provides battlefield digital data management technologies that are designed to save users time and money in collecting, storing, and securing their data assets. The major component of this process is Deep Archive® software (DA). DA is a client-based application that takes digital data and

compresses, encrypts, metadata tags, and sends to various Storage Area Network (SAN) assets existing with a customer. The philosophy that Deep Archive software employs is simple: Before any data ever leave a desktop anywhere globally, either from the Radiology Tech, or from a surgeon sending local pictures back to his colleagues for consultation, that data is compressed, encrypted, metadata tagged, and then sent. Because the file is already smaller before it enters the data network, it travels more efficiently (saving time and money), and it is stored using less space and at lower costs. Server-based compression, while effective in some circumstances, requires large network "pipes" just to get the image to the server.

CONTACT NAME: Glen Moore
E-MAIL: gmoore@deeparchive.com
WEBSITE: http://deeparchive.com

Ekagra Software Technologies

A licensed caBIG®Support Service Provider

Ekagra is a well established small business with extensive experience with federal programs, specializing in health IT, bioinformatics, open source technologies, and information management in support of scientific research and healthcare. We provide a disciplined, engineering-based approach to all of our service engagements to deliver consistent quality results. Ekagra provides full life cycle information technology services from project planning and enterprise architecture to application development and deployment. Ekagra is a NCI-licensed caBIG® Support Service Provider for Adaptation and Enhancement of caBIG®-compatible Software Applications and for deployment support for caBIG® Applications. Our staff members are experts in designing service-oriented architecture and integration of legacy systems. We assist customers in all their needs from development and adaptation to adoption and deployments. We offer hosting, cloud computing, and on-site support services. Ekagra recognizes that information technology is rapidly evolving and customers will require partners that are equally agile. Ekagra has demonstrated experience in our ability to respond quickly and adapt to the needs of our customers. Ekagra has been an active participant in and key contributor to the caBIG® program since its inception. As IT leaders in this domain, our SMEs have a respected voice that is recognized in the community. Ekagra has experience with caBIG® tools, technologies, and communities. This experience provides a cadre of personnel to guide future tasks, greatly reducing the ramp up time of new teams; provides a real guidance capability at the program level; and helps with the coordination of our team with other projects. Our broad experience in the caBIG® domain in the ECCF, sSOA, caCORE infrastructure, caGrid, metadata management, clinical trials, and life sciences data integration allows us to apply our engineering skills to solve the business problems in health research and care. Our deep and broad expertise lowers risk, reduces learning costs, allow much faster start up (as demonstrated in a caGrid deployment task in 2008 where we helped deploy caGrid in 30+ Cancer centers). We bring a track record of focused delivery for large and complex federal programs. We use engineering-driven methods and industry standard processes and tools to plan and deliver work. Our processes encourage active engagement of stakeholders and customers to ensure predictable schedule and results.

CONTACT NAME: Kalpesh Patel
E-MAIL: kpatel@ekagrasoft.com
WEBSITE: www.ekagrasoft.com

Elekta Impac Software

Elekta's human-centric software creates an efficient clinical environment in which all activities related to patient care—from diagnosis and treatment to follow-up—are as streamlined as possible, giving clinicians more time to focus on patients. Elekta's open systems and vendor-neutral connectivity ensure cross-platform flexibility to integrate the most advanced and useful tools. Oncology departments worldwide recognize Elekta Software's unique, comprehensive offerings, with 100,000 patients annually receiving diagnosis, treatment, or follow-up facilitated by Elekta solutions. Elekta has an oncology-specific electronic medical record MOSAIQ that connects to the Coalition of Cancer Group's TrialCheck database of oncology clinical trials via a caBIG® Bronze-certified interface.

CONTACT NAME: Marjorie van der Pas
E-MAIL: mvanderpas@impac.com
WEBSITE: www.elekta.com/

ESAC, Inc.

A licensed caBIG® Support Service Provider

Enterprise Solutions and Consulting Incorporated (ESAC, Inc.) is an 8(a) certified small business delivering project management and technology solutions to government, private industry, and academia. A caBIG® -licensed Support Service Provider in the category of Adaptation and Enhancement of caBIG®-Compatible Software Applications, ESAC, Inc., brings subject matter expertise, quality assurance, and project management services to caIntegrator, caB2B, caGWAS, and caTissue. From providing custom application development, data validation, bioinformatics, and quality assurance at the bench to contributing data management and quality assurance expertise to large-scale projects such as the Cancer Genome Atlas Project, ESAC has joined with diverse teams to produce bioinformatic solutions that work. ESAC, Inc., has a thorough knowledge of the caBIG®-compatibility review process and can adapt existing systems to utilize caBIG® services and APIs. We are an experienced systems integrator and can work with your institution's research staff to integrate caBIG® compatible systems into your research and development workflows. After success deploying the caTissue Suite at the Georgetown Lombardi Comprehensive Cancer, ESAC, Inc., was selected to support the Georgetown Database of Cancer, a clinical decision tool utilizing the latest discoveries in molecular medicine. ESAC, Inc., is currently involved in the design and development of several tools, including caIntegrator for the NCI CBIIT caBIG® Integrative Cancer Research workspace. In addition, we provide software design and development to the Informatics Support Center for the Breast and Colon Cancer Family Registry program at the NCI Division of Cancer Control and Population Sciences. For projects that encompass a range of technical and biological challenges, the experience and commitment of ESAC, Inc., can help you bring them to life.

CONTACT NAME: Henry Schaefer
E-MAIL: henry.schaefer@esacinc.com
WEBSITE: http://esacinc.com

Georgetown University Lombardi Comprehensive Cancer Center

The Informatics Group at Lombardi Comprehensive Cancer Center has a diverse and expansive portfolio of ongoing projects relevant to the caBIG® program. The caBIG® In Silico Research Center of Excellence (ISRCE): Our intent is to utilize caBIG® tools and resources to integrate data across multiple "omics" platforms, along with detailed clinical information, to develop novel data analytic methods. Our goal is allow medical professionals to stratify patients in order to optimize therapy and monitor recurrence in cancer patients at the Lombardi Cancer Center. During the

course of this effort, new caBIG®-compatible data and analytic grid services will be created, including a small molecule library for identified targets. However, the most promising products currently in development involve the application of machine learning methods, including Phenotypic Up-regulated Gene Support Vector Machine (PUG-OVRSM) and Differential Dependency Network (DDN), to identify new drug therapies for tamoxifen-resistant breast cancers. Opportunities for collaboration exist. The Georgetown Database of Cancer (G-DOC): This tool is designed to serve as a cutting-edge data integration platform and integrative knowledge discovery system for the oncology and translational research communities. By aggregating public and proprietary clinical and "omics" data from across the Medical Center, G-DOC is expected to help bring about significant advances in personalized medicine for patients and to promote identification of new drug targets and therapeutic modalities. Research prototypes developed in the caBIG® ISRCE program at Georgetown are productized and integrated into this platform to their ensure accessibility and ease of use for the community. G-DOC also integrates caTissue. Cancer Bench to Bedside (caB2B): The Informatics Team at Georgetown University engaged in ongoing work to develop the next generation caB2B. This process involves careful solicitation of stakeholders needs to identify a variety of translational research use cases, which will then be triaged with NCI for implementation prioritization. This subject is covered in more detail in a separate poster. Those wishing to contribute to next-generation requirements gathering are welcome to get involved.

CONTACT NAME: David Tanenbaum
E-MAIL: dmt45@georgetown.edu
WEBSITE: https://informatics.lombardi.georgetown.edu

Healthcare IT, Inc.

A licensed caBIG® Support Service Provider

HealthCare IT, Inc. (HCIT) is a small business where technology meets science and medicine. HCIT is comprised of information technology professionals who understand the clinical research and healthcare industry. We proudly serve the cancer research community as a caBIG® Support Service Provider in the category of Deployment Support for caBIG® Software Applications, and offer hosting capabilities for caBIG® applications under a Software-as-a-Service model. HCIT has re-engineered the open source code for caMatch to assist in the collection and analysis of information from consenting breast cancer patients. In a hosting capacity, HCIT is working with the NCI on the current Patient Outcomes Data Service (PODS). Additionally, HCIT is working with the early release of the Common Biorepository Model (CBM) for integration with our commercial biobanking management product known as BIGR®. As the flagship product and service, the BIGR® biospecimen repository management system is the most comprehensive yet customizable software product in the biospecimen space. BIGR® takes a personalized approach while allowing BioBanks and Biospecimen Repositories to enter, track, and manage all aspects of sample management to their own specifications while adhering to federal, state, and internal policies and standards. The Healthcare IT, Inc., data center team provides reliable and secure managed hosting services for many other healthcare and clinical research applications that allow our customers to concentrate on what they do best—making discoveries that save lives and make major impacts in the life sciences community.

CONTACT NAME: Lisa Locascio
E-MAIL: llocascio@healthcit.com
WEBSITE: http://healthcit.com

Information Management Services, Inc.

A licensed caBIG® Support Service Provider

IMS has extensive experience with the caBIG® Workspaces (WS) and Special Interest Groups (SIGs). IMS has been an active participant in these caBIG® activities since the onset of the initiative. IMS has assigned senior staff to participate in the caBIG® Workspaces and SIGs appropriate to their area of expertise. This involvement has consisted of participation in group conference calls, face-to-face meetings, training opportunities, and as a caBIG® participant and exhibitor in the caBIG® annual meetings. In this capacity, IMS participated in the addition of the NAACCR data elements to the caDSR so that cancer registries can build caBIG® compliant tools. The NAACCR data elements in caDSR provides a mechanism for the IMS-developed SEER*DMS software to allow data sharing using caBIG® data standards. In addition, as a participant in the TBPT workspace, IMS has been involved with the development of the Common Biorepository Model. IMS designs and develops systems for the collection, storage, retrieval, and analysis of cancer information as well as systems for the administration of cancer research. IMS has also developed a biological specimen repository management system (BSI). IMS has a BSI web services application that is certified as caBIG® Bronze campatible and implemented the CBM into BSI. The BSI application can extract and publish specimen data using caBIG® CDEs. IMS also developed the caSEER system under the NCI/DCCPS SEER Program as a method to provide cancer surveillance information via the caGrid. caSEER involved the deployment of seven databases and an analytic application, which queries these databases and visualizes the results on the grid. caSEER provides cancer incidence, mortality, behavior risk factor, population demographics, and DHHS Healthy People 2010 Goals analyses from various SEER public use datasets. Resultant statistics are presented in the form of tables, charts, and geographic maps. In addition, the BSI web services and the caSEER application comply with the caBIG® Silver compatibility guidelines. IMS is a licensed caBIG® Support Service Provider, in the category of Adaptation and Enhancement of caBIG®-Compatible Software Applications. IMS utilizes the develop methodology and tools associated with the caCORE Software Development Kit to extend existing caBIG® compatible software products. New object models and common data elements are also developed utilizing CBIIT's repository, the Cancer Data Standards Repository (caDSR). The Common Security Model (CSM) and caAdapter are utilized as required by the particular application or tool.

CONTACT NAME: Janis Beach
E-MAIL: beachj@imsweb.com
WEBSITE: http://imsweb.com

INFOTECH Soft, Inc.

A licensed caBIG® Support Service Provider

INFOTECH Soft is licensed by the National Cancer Institute as a caBIG® Support Service Provider in the category of Adaptation and Enhancement of caBIG®-Compatible Software Applications. We are dedicated to the development of cutting-edge, intelligent, flexible knowledge discovery software solutions for cancer and other biomedical research.
Our ontology-based GeneTegra system enables concept-based integrated querying of data sources from caGrid and beyond. It provides a graphical user interface to explore and search ontology models of caBIG® semantics, to integrate caGrid services with other external sources, and to create and execute queries against the concepts defined in these representations. Our experienced team of researchers and technical staff applies the latest developments in the Semantic Web and a deep understanding of the caBIG® technology stack to create, adapt, and enhance innovative software solutions. We have provided development and adaptation

services to the Center for the Development of a Virtual Tumor (CViT.org) at Massachusetts General Hospital (MGH) for the creation and caGrid-enablement of the Digital Model Repository (DMR), for the development of the Computational Model Execution Framework to deploy and execute DMR models, and for the creation of a transatlantic environment that integrates the DMR with European repositories.

CONTACT NAME: Patrick Shironoshita
E-MAIL: patrick@infotechsoft.com
WEBSITE: www.infotechsoft.com

LabAnswer

A licensed caBIG® Support Service Provider

LabAnswer specializes exclusively in laboratory and clinical informatics consulting. The focus of our work is largely FDA-regulated industries involving bioinformatics in all phases of clinical trials (including cancer research), vaccine development, and bio-banking. We also do significant laboratory automation work in pharmaceutical manufacturing, medical research (including cancer and functionality offered by caBIG® tools), medical device, public health, and forensics laboratories. LabAnswer has the ability to architect, design, configure, develop, implement, deploy, and support nearly any laboratory informatics application (custom, open source, and COTS) based on our domain expertise and skill-sets. Most of our employees are scientists/laboratory professionals with senior-level IT experiences and capabilities. Our depth and breadth of experiences working with laboratory informatics applications is unmatched by any other company or group of companies. LabAnswer is a licensed caBIG® Support Service Provider by the National Cancer Institute (NCI) in the categories of Help Desk Support, Deployment Support for caBIG®-Compatible Software Applications, and Documentation, Training Materials, and Services.

CONTACT NAME: Jim Goodmiller
E-MAIL: jim.goodmiller@labanswer.com
WEBSITE: www.labanswer.com

LabVantage

One of the key issues facing organizations today is the complexity of managing the increasing number of bio-specimens (tissue, blood, DNA, RNA, etc.) generated during the research process. With heightened regulatory and privacy compliance necessitated by Good Laboratory Practices (GLPs), Good Clinical Practices (GCPs), 21 CFR Part 11, Health Insurance Portability and Accountability Act (HIPAA) and Internal Review Board (IRB) requirements, a solution that accurately tracks patient consent, HIPAA waivers, laboratory conditions, and electronic signatures is critical. Furthermore, as organizations recognize the additional value of utilizing samples during the discovery process, solutions need to be able to easily store, locate, and access vital biospecimens along with critical genomic, proteomic, and phenotypic information without difficult system integration. Recognizing the diverse challenges of running a successful and efficient biorepository, LabVantage has created the SAPPHIRE Biobanking Solution. Developed in conjunction with the original developer of the Spectrum™ specimen management system, the biobanking module addresses the need for data capture, location management, handling assurance, and operational efficiency required to manage a biorepository at any organization. SAPPHIRE enables centralized biological specimen tracking from “cradle to grave" to comply with GLPs and GCPs, satisfy patient consent requirements, enhance scientific accuracy, and improve development efficiency. Whether storing or tracking whole-blood, tissue, cellular lysates, DNA, RNA, proteins, etc., SAPPHIRE supports organization-wide inventory control. SAPPHIRE offers intricate chain-of-custody functionality, including detailed location and shipment management, aliquot/derivative

and pooled sample tracking, and electronic signature captured transfer and disposition. Moreover, as an integrated solution, SAPPHIRE enables organizations to track the genomic and phenotypic data associated with these biospecimens under compliance with HIPAA and other privacy laws. LabVantage's SAPPHIRE is certified as caBIG® Bronze compatible, and we provide repository inventory management and LIMS solutions.

CONTACT NAME: Mark Cada
E-MAIL: mcada@labvantage.com
WEBSITE: www.labvantage.com

Medidata Solutions

Medidata Solutions (www.mdsol.com) is a leading global provider of clinical technology solutions that enhance the efficiency of customers' clinical development processes and optimize their research investments. Medidata products and services allow customers to achieve clinical results more efficiently and effectively by streamlining the design, planning, and management of key aspects of the clinical research processes, including protocol development (Medidata Designer®) and the capture, management, analysis, and reporting of clinical trial data (Medidata Rave®). Medidata's diverse customer base spans Government agencies, academic and not for profit institutions, pharmaceutical, biotechnology and medical device companies, CROs, and other research organizations, and includes the National Cancer Institute, City of Hope, National Cancer Institute of Canada, and the Ludwig Institute for Cancer Research. NCI has purchased licensing rights to use and distribute the Medidata Rave Electronic Data Capture / Clinical Data Management System (CDMS) software package throughout the NCI Clinical Research Enterprise. This includes but is not limited to, NCI-designated Cancer Centers; NCI Clinical Trials Cooperative Groups; research organizations that the Institute funds as Specialized Programs for Research Excellence (SPOREs); NCI Community Clinical Oncology Programs (CCOPs) and Minority-Based Community Clinical Oncology Programs (MBCCOPs); and components of the National Community Cancer Centers Program (NCCCP).

CONTACT NAME: Gregg Kravatz
E-MAIL: gkravatz@mdsol.com
WEBSITE: www.mdsol.com

Moxie Informatics

A licensed caBIG® Support Service Provider

Moxie Informatics is a health informatics company comprised of experts who are skilled and experienced in the business of life science research and health care. Our team is intimately familiar with the business of the National Cancer Institute (NCI) and many of its offices and programs. Moxie's team resources currently interface with NCI Directors, Program Managers, and other government or contractor experts who are working within the CBIIT/caBIG® enterprise. Our specific focus is providing innovative value-added information technology services to the scientific research community. Moxie Informatics' mission is to harness the power of information technology to help advance scientific research. We build caBIG®-compliant informatics applications that deliver a high return on investment to our clients. We intend to emerge as a trusted provider of informatics systems. Moxie Informatics is a licensed caBIG® Support Service Provider in the category of Adaptation and Enhancement of caBIG®-Compatible Software Applications.

CONTACT NAME: Michael Connolly
E-MAIL: connollym@moxieinformatics.com
WEBSITE: http://moxieinformatics.com

NCRI Informatics Initiative

The main role of the NCRI Informatics Initiative is to promote the technological and community cultural advances required to achieve our objectives of researchers being able to share data and tools seamlessly for

the benefit of scientific advancement and ultimately patients. The central core of our work is to liaise with the UK research community to ensure that data generated across different organisations becomes available for others to use. That entails not only encouraging a willingness to publish and share such data, but also the promotion of the existence of such data, and the technical aspects required to make it easy to integrate the data with other resources. Many of the underlying issues are national and international in nature, and a key role of the Initiative is to work with our strategic partners to encourage the use of standardised datasets and exchange mechanisms that will eventually allow the linking of different research datasets on a national, and ultimately global, basis.

CONTACT NAME: Stuart Bell
E-MAIL: stuart.bell@ncri.org.uk
WEBSITE: www.cancerinformatics.org.uk

Persistent Systems

A licensed caBIG® Support Service Provider

Persistent Systems is recognized as an award-winning technology company specializing in software product development services. We have delivered more than 2,400 software product releases to our 200+ customers in the last six years. We have also developed reusable assets and frameworks, and proven processes for the entire software product lifecycle that reduces time to market while delivering consistent quality, as evidenced by customer partnerships that span an average of five years. Life Sciences and Health Care is our second-largest vertical practice, and we have developed and deployed solutions for Clinical informatics and Translational research, Bioinformatics, and Lab Automation. Our customers range from academic institutions, cancer centers, instrumentation and medical device vendors, to large pharmaceutical and biotech companies. Our Life Sciences team has also been involved in the caBIG® initiative from its inception and we are a licensed caBIG® Support Service Provider. Our technology offerings include software development, adoption, deployment, grid enablement, SaaS/Cloud enablement, data management/BI, and support services.

CONTACT NAME: Aditya Phatak
E-MAIL: aditya_phatak@persistentsys.com
WEBSITE: http://www.persistentsys.com

Recombinant Data Corporation

A licensed caBIG® Support Service Provider

Recombinant Data Corp. is a licensed cancer Biomedical Informatics Grid® (caBIG®) Support Service Provider in the category of deployment support for software applications in the caBIG® Life Sciences Distribution (LSD). Recombinant provides leading-edge data warehousing and clinical intelligence solutions to healthcare providers, academic medical centers, and life science researchers to deliver higher quality outcomes, accelerate personalized medicine, and lower costs. Our team of industry veterans is focused on improving the flow of reliable data to power clinical and research applications in a secure, compliant environment.

CONTACT NAME: William Stetson
E-MAIL: wstetson@recomdata.com
WEBSITE: www.recomdata.com

RURO, Inc.

RURO, Inc., headquartered in Frederick, Maryland, specializes in development and production of state-of-the-art computer software for research, biotechnological, pharmaceutical, healthcare, and government (homeland security) laboratories in the U.S. and worldwide. Our recent line of biological applications is designed to increase the productivity in the labs while maintaining the highest level of security, versatility, and knowledge. Our product FreezerPro allows scientists and managers in research laboratories and pharmaceutical companies to inventory and

retrieve frozen laboratory samples, and we are certified as caBIG® Bronze compatible.

CONTACT NAME: Vlad Lebedev
E-MAIL: info@ruro.com
WEBSITE: www.ruro.com

SAFE-BioPharma Association

SAFE-BioPharma Association is the non-profit association that created and manages the SAFE-BioPharma digital identity and signature standard for the pharmaceutical and healthcare industries. The SAFE-BioPharma standard:

- mitigates legal, regulatory, and other business risk associated with electronic transactions;
- facilitates interoperability between disparate information systems;
- provides a secure, enforceable, and regulatory-compliant way to verify identities and apply digital signatures in electronic transactions; and
- helps green the pharmaceutical and healthcare industries by dispensing with paper originals and other cumbersome forms of back up.

SAFE-BioPharma is a global, interoperable digital identity and signature standard for use across the biopharmaceutical and healthcare industries. Others, such as VeriSign, are individual vendors with specific product biases. As a standard, SAFE-BioPharma is technology and vendor neutral and avoids the complexity of dealing with multiple digital signatures across multiple trust networks. SAFE-BioPharma also works closely with regulatory agencies to ensure they understand and accept SAFE-BioPharma digital identities and signatures. We are currently in the process of doing a pilot project with the NCI regarding identity authentication and digital signatures.

CONTACT NAME: Tanya Newton
E-MAIL: tanya.newton@safe-biopharma.org
WEBSITE: www.safe-biopharma.org

SAIC

A licensed caBIG® Support Service Provider

Science Applications International Corporation (SAIC) is a leading provider of scientific, engineering, systems integration, and technical services and solutions to all branches of the Department of Health and Human Services, U.S. military, agencies of the Department of Defense (DoD), the Intelligence Community, the U.S. Department of Homeland Security, and other U.S. government civil agencies, as well as to customers in selected commercial markets. With more than 44,000 employees in more than 150 cities worldwide, SAIC engineers and scientists solve complex technical challenges requiring innovative solutions for customers' mis-sion-critical functions. SAIC performs contract research and development for some of the leading R&D-funding agencies in the United States, including the National Institutes of Health (NIH), National Cancer Institute (NCI), Defense Advanced Research Projects Agency (DARPA), Department of Energy (DOE) National Labs, and DoD laboratories, as well as major commercial companies in the life sciences field. SAIC has successfully provided bioinformatics support to the NCI Center for Biomedical Informatics and Information Technology (CBIIT) since 2000.

CONTACT NAME: Susan Foster
E-MAIL: susan.a.foster@saic.com
WEBSITE: www.saic.com/health

Sapient Government Services, Inc.

A licensed caBIG® Support Service Provider

Sapient delivers full lifecycle software engineering solutions and services to the federal government, including software engineering, database design, implementation, administration, data quality assurance and Independent Verification & Validation (IV&V), build and deployment automation, human factors testing, technical writing, and development of training

materials. We employ industry best practices, and our extensive experience in developing applications for the scientific community helps ensure that the technical and system designs operate according to user, performance, and usability requirements. Our health practice is one of our most robust both within the commercial and government sectors. For healthcare and related institutions and companies, the Sapient team has provided program management, design, enterprise architecture, development, data center services, security, validation, and implementation. Our focus has been on: (1) connecting researchers and clinicians through a shareable and interoperable infrastructure; (2) developing standard rules and a common language to share information more easily; and (3) building and adapting tools for collecting, analyzing, integrating, and disseminating information associated with health research and care. Sapient Government Services is headquartered in Arlington, Virginia. Sapient is a leading provider of consulting, technology, and marketing services. We help our clients transform their technologies, systems, and processes using our knowledge of commercial best practices coupled with expertise in the federal government and health services infrastructure and architecture. Sapient is a licensed caBIG® Support Service Provider in the category of Adaptation and Enhancement of caBIG®-Compatible Software Applications.

CONTACT NAME: Stephen A. Goldstein
E-MAIL: sgoldstein@sapient.com
WEBSITE: www.sapient.com/

ScenPro, Inc.

A licensed caBIG® Support Service Provider

ScenPro has worked with the National Cancer Institute continuously since 1997, supporting the institute's goals through domain modeling, enterprise system development, software application development, training, and long-range planning support. We are a licensed caBIG® Support Service Provider in the areas of Documentation and Training Materials and Services, as well as Adaptation and Enhancement of caBIG®-Compatible Software Applications. As such, we are committed to supporting organization's efforts in achieving the overall caBIG® program goal of semantic interoperability. Using our extensive knowledge and expertise of caBIG® tools, technologies, standards, and development protocols, we are able to provide dynamic relevant services to those interested in adopting caBIG® applications or adapting existing healthcare applications for caBIG® interoperability. Our knowledge of the caBIG® enterprise, familiarity with NCI CBIIT development best practices, and strong relationships with project managers, development teams, trainers, documentation team members, end users, and subject matter experts both within NCI CBIIT and across the caBIG® community give us the ability to provide you with the customized services required to achieve your caBIG® vision.

CONTACT NAME: Charles Yaghmour
E-MAIL: cyaghmour@scenpro.com
WEBSITE: www.scenpro.com

SemanticBits

A licensed caBIG® Support Service Provider

SemanticBits specializes in the design and development of software systems for the health and life sciences industry with a specific focus on:

- using open source technology;
- reusing existing infrastructure to help minimize development cost;
- working in multi-disciplinary environments; and
- staying committed to serving the individualized needs of our customers.

In order to achieve the best quality, we practice an iterative, incremental development process that tackles high-risk,

challenging tasks first and emphasizes testing. We strongly believe in taking a domain-driven approach, where we place an emphasis on the requirements of our clients in the context of their domain. Our 10+ years of experience in biomedical informatics provides us with the necessary background to tackle the complex problems of the clinical and bioinformatics fields. SemanticBits has the resources and experience to tackle virtually any aspect of a project, including software engineering, project management, business analysis, life sciences subject matter expertise, quality assurance, engineering, usability engineering, and technical writing.

CONTACT NAME: Zainab Zaki
E-MAIL: zainab.zaki@semanticbits.com
WEBSITE: www.semanticbits.com

Share Your caBIG® Story

The caBIG® program has always been a collaborative initiative, where experiences and best practices are shared broadly for the benefit of the community. Whether you are just getting started or have been part of the caBIG® community for years, we want to hear about your experience. The purpose of the *Share Your Story* campaign is to learn about the community's experience with the caBIG® initiative at all levels of engagement—from technology evaluation and decision-making to development, implementation, or use of caBIG®-interoperable technology. If you are just getting started, tell us what research challenges you are trying to address. If you are in the process of developing or deploying caBIG® technologies including software, caGrid, semantics, and data-sharing policies, we'd love to hear how this is unfolding in the real world for you and your colleagues. If, on the other hand, you are a long-time veteran, please share the latest ways that caBIG® capabilities are being used at your organization to change the way discovery and translational research are being conducted, clinical care is delivered, or data liquidity is becoming a reality. Representatives will be onsite at the *Share Your Story* booth ready to hear from you about challenges and successes as you move along the path to caBIG® interoperability. With your permission, information gathered onsite may be used to develop case studies and materials to help future generations of caBIG® participants as they leverage caBIG®-interoperable technology to solve biomedical research challenges. Participating in this initiative will also allow others in the biomedical community to learn about your work. Stop by and see us!

CONTACT NAME: caBIG® Program Team
E-MAIL: http://cabig.cancer.gov/action/shareyourstory
WEBSITE: http://cabig.cancer.gov

Siemens Corporate Research

A licensed caBIG® Support Service Provider

Siemens Corporate Research (SCR) is a licensed caBIG® Support Service Provider in the category of Adaptation and Enhancement of caBIG®-Compatible Software Applications. For thirty-three years Siemens Corporate Research (SCR) has been engaged in software research, engineering, and consulting. Our staff of approximately 300 researchers in Princeton, New Jersey, has brought a wealth of software and systems engineering experience to our research, development, and consulting projects. Siemens Corporate Research is part of the Siemens Corporation in the U.S., which is part of Siemens AG, a large multinational electronics and electrical engineering company established more than 160 years ago in Germany by Werner von Siemens. Research and Development (R&D) groups within SCR innovate in a broad range of application areas. Among these are: (1) *Imaging and Visualization*, including computer imaging and visualization applied in medicine; content extraction and object modeling; 3-D interactive processing and imaging architecture; interventional imaging; and interventional MR; (2) *Integrated Data Systems*, including data and information

integration; structuring and modeling; analysis, understanding, and personalization; knowledge management; robust analysis and content retrieval; personalized healthcare; data integration; modeling and optimization; intelligent data analysis; and knowledge management; and (3) *Real-time Vision and Modeling*, including real-time vision for safety, security, transportation, automation, automotive, maintenance, and service; smart cameras and machine vision; 3-D vision and augmented reality; statistical methods for vision, audio, and signal processing; and adaptive techniques. SCR has a caBIG® team comprising members from these different R&D groups. Together they work on projects such as the eXtensible Imaging Platform (XIP) and Algorithm Verification Tool (AVT).

CONTACT NAME: Daniel J. Paulish
E-MAIL: daniel.paulish@siemens.com
WEBSITE: www.usa.siemens.com

Terrapin Systems® LLC (Terpsys)

A licensed caBIG® Support Service Provider

Terrapin Systems® LLC is a customer service company focused on technology, offering licensed caBIG® Support Services in three (3) separate areas:

- Our caBIG® Training Materials and Services include on-or-off-site caBIG®-compatible applications development and caCORE infrastructure support.
- Our caBIG® Deployment Support Services feature customized and standard server platform builds, configurations, installations, upgrades, and troubleshooting.
- Our comprehensive caBIG® Help Desk Services provide 24x7 phone and online support, as well as incident tracking and communications plans for system-wide outages and emergencies.

CONTACT NAME: Craig Fee
E-MAIL: cfee@terpsys.com
WEBSITE: www.terpsys.com/

The Ohio State University

The Ohio State University Center for IT Innovations in Healthcare (CITIH) engages in research, development, and evaluation activities intended to yield innovative IT informatics platforms for use by the healthcare and biomedical sciences community. CITIH's fundamental mission is to seek innovative IT solutions to problems in biomedical science and healthcare delivery through collaboration with practitioners, researchers, external academic medical centers, and industry partners. A common principle across all CITIH research, development, and evaluation programs is the active pursuit of technology and knowledge transfer between center members and our collaborating external academic institutions and industry partners. A foundation of CITIH's collaboration with external partners is the incorporation of multidisciplinary team-based approaches spanning such public-private partnerships during all phases of the center's projects and initiatives.

CONTACT NAME: Tome Nielsen
E-MAIL: tome.nielsen@osumc.edu
WEBSITE: http://citih.osumc.edu/

Velos, Inc.

Velos is the trusted clinical trial management resource for investigators, sponsors, and academic leaders throughout the U.S. Velos eResearch, which is certified as caBIG® Bronze compatible, is deployed for clinical research in all parts of the healthcare enterprise and supports a broad diversity of clinical departments and functional needs. The system fundamentally improves the way data is collected, organized, and shared. A pure Internet technology platform, Velos eResearch harnesses the Internet to enable research

sites, sponsors, and patients to participate in a secure, integrated platform. System users are freed of redundant data entry and related time delays associated with most clinical research today. With emphasis on workflow, Velos integrates the clinical, administrative, and financial information needs of research management. In the cancer community, eight of the top twenty cancer centers are Velos eResearch customers and 40% of NCI extramural funding to the top 100 such recipients goes to Velos customers, the largest such market share among clinical trial management and electronic data capture systems. Velos is privately held with headquarters in Fremont, California.

CONTACT NAME: Michele McKenna
E-MAIL: bwilkerson@velos.com
WEBSITE: www.velos.com

Westat

Westat, located in Rockville, Maryland, is one of the foremost contract research corporations in the United States. We conduct surveys and provide research and related services to the agencies of the U.S. Government and to a broad range of institutional and business clients. We are a rapidly growing employee-owned corporation with over 2,000 research, technical, and administrative staff and a forty-five-year history of research and managerial excellence. Westat provides a wide range of research services in the areas of cancer bioinformatics and population-based cancer studies. Our cancer bioinformatics experience spans the development life cycle of a protocol, from protocol design, to forms and database design, to analyses of trial results. Our population-based cancer research experience includes data linkages at the cohort, registry, or population level; Geographic Information Systems (GIS) analyses; health information coding, data abstracting, and management; and vital records management. Westat provides an array of data management services and solutions, including clinical database design, data collection, data quality management, and operational study reports. As part of our data management solutions, Westat offers the Biological and Environmental Sample Tracking (BEST) Software System. BEST is a comprehensive web-based solution to manage the lifecycle of biospecimens and environmental samples. BEST is easily configured so that protocol-specific collection, storage, and shipment requirements, including workflow and validation, can be defined and ready for use in a short period of time. Westat also offers the Data Delivery Metadata System (DDMS), which provides an automated and structured approach to the management of data deliveries and associated metadata (such as data dictionaries, codebooks, annotated instruments, and decision logs). The DDMS provides a standardized tracking of multiple releases of datasets during data collection; a simple way to create and maintain detailed decision logs on data corrections; cross-links for metadata review and validation; and an infrastructure that will allow for quick and efficient response to data enquiries by analysts and investigators.

CONTACT NAME: Keith Umbel
E-MAIL: keithumbel@westat.com
WEBSITE: www.westat.com

4c. ESN

The caBIG® Enterprise Support Network (ESN)

The caBIG® initiative offers a number of support resources to meet the diverse needs of those learning about and deploying caBIG® technologies and resources. Chief among these is the Enterprise Support Network, including both Knowledge Centers and Support Service Providers.

Knowledge Centers: Domain Expertise

Six caBIG® Knowledge Centers (KCs) have been established at institutions with demonstrated expertise in a specific area of focus or domain of interest. They serve as the nexus for an expanding community employing caBIG® tools, standards, and infrastructure in a specific domain. KC staff can provide expert guidance to end-users, IT staff, and senior decision makers implementing caBIG® tools and infrastructure.

- **caGrid**: Led by The Ohio State University and The Ohio State Comprehensive Cancer Center, with the University of Chicago and the Argonne National Lab, the caGrid KC provides knowledge resources for those interested in learning about, using, and contributing to caGrid, the backbone of the caBIG® infrastructure.
- **Clinical Trials Management Systems**: Led by Duke University Comprehensive Cancer Center, with the Robert H. Lurie Comprehensive Cancer Center at Northwestern University, Cancer, Leukemia Group B – Information Systems (CALGB-IS), and SemanticBits, the CTMS KC provides knowledge resources related to clinical trials management, including stewardship of the following caBIG® tools: Cancer Center Participant Registry (C3PR), Cancer Adverse Events Reporting System (caAERS), Patient Study Calendar (PSC), Integration Hub, Labviewer, and Clinical Connector.
- **Data Sharing and Intellectual Capital**: Led by the University of Michigan, the DSIC KC provides centralized, authoritative repository of processes, model agreements, and other resources to encourage and facilitate data sharing to advance scientific discovery consistent with applicable legal, regulatory, ethical, and contractual requirements.
- **Molecular Analysis Tools**: Led by Columbia University Herbert Irving Comprehensive Cancer Center with The Broad Institute of MIT and Harvard, the MAT KC provides knowledge resources related to molecular analysis, including stewardship of the following caBIG® tools: caArray, caIntegrator, geWorkbench, and GenePattern.
- **Tissue/Biospecimen Banking and Technology Tools**: Led by the Siteman Cancer Center, Washington University at St. Louis, this KC provides knowledge resources related to biospecimen management, including stewardship of the following caBIG® tools: caTissue Suite, caTissue Core, and caTIES.
- **Vocabulary**: Led by Mayo Clinic, the Vocabulary KC provides knowledge resources for individuals and institutions interested in making use of or extending caBIG® and other vocabulary tools and provides stewardship of the following tools: LexBIG/LexEVS, NCI Protégé, and LexWiki.

Each KC will provide more information about the expertise and range of services it offers at tables located throughout the World's Fair.

Support Service Providers: Comprehensive Technical Support

caBIG® Support Service Providers (SSPs) are independent organizations that are approved by NCI as meeting specific criteria for performance. With licenses in one or more of four specific service categories, SSPs provide client-specific caBIG® support under negotiated business arrangements with the client independent of the NCI and the caBIG® program.

Many SSPs have signed up to exhibit at the annual meeting and will provide more information about the range of services they are licensed to offer during the World' Fair. Some will also deliver presentations and answer questions during SSP Vendor Theater (see the schedule and participating organizations which is included in this Tab section).

SSPs attending Annual Meeting:

Akaza Research

5AM Solutions

CTIS, Inc.

Ekagra Software Technologies

E-SAC, Inc.

HealthCare IT, Inc.

Information Management Services, Inc.

INFOTECH Soft, Inc.

LabAnswer

Moxie Informatics

Persistent Systems

Recombinant Data Corporation

SAIC

Sapient

ScenPro, Inc.

SemanticBits

Siemens

TerpSys

Help Desk Support

This service category includes support for end users and local IT administrators using the caBIG® applications and tools contained within the three caBIG® bundles (Life Sciences Distribution, Clinical Trials Compatibility Framework, and Data Sharing and Security Framework), as well as those wishing to adapt tools or connect their tools to the grid.

- Asclepius Solutions
- CTIS, Inc. (Clinical Trials Compatibility Framework)
- LabAnswer
- Persistent Systems
- TerpSys

Adaptation and Enhancement of caBIG®-compatible Software Applications

This service category supports caBIG®-compatible software development and includes adaptation of existing systems for caBIG® compatibility, custom enhancements to existing caBIG® tools that satisfy specific user-driven requirements while maintaining caBIG® compatibility, and de novo development of caBIG®-compatible applications and tools.

- 5AM Solutions
- ESAC, Inc.
- Ekagra Software Technologies
- Information Management Services, Inc.
- INFOTECH Soft, Inc.
- Moxie Informatics
- Persistent Systems
- SAIC
- Sapient Government Services, Inc.
- ScenPro, Inc.
- SemanticBits
- Siemens Corporate Research

Deployment Support for caBIG® Software Applications

This service category includes onsite or offsite procurement and provisioning of hardware, operating systems and other software, such as application servers and databases, along with installation and configuration of caBIG® software and legacy data transformation and migration. Deployment support also includes integration of multiple caBIG® tools and caBIG®-compatible applications through open application programming interfaces (APIs). Other possible services include data management, security service, hosting, performance tuning, network connectivity, grid enablement, and maintenance of Web server/database and application-specific administrative IT tasks.

- 5AM Solutions
- Akaza Research
- Asclepius Solutions
- CTIS, Inc.
- Ekagra Software Technologies
- Healthcare IT, Inc.
- LabAnswer
- Persistent Systems
- Recombinant Data (Life Sciences Tools)
- SemanticBits
- SRA Corporation
- Terpsys

Documentation and Training Materials and Services

This service category includes development of documentation and training materials and the provision of training services for caBIG® software applications, caBIG® compatibility for existing systems, and use of caCORE infrastructure.

- Akaza Research (Training Materials and Services)
- CTIS, Inc.(Training Materials and Services)
- LabAnswer
- Persistent Systems
- ScenPro, Inc.

- SemanticBits (Documentation Materials and Services)
- Terpsys (Training Materials and Services)
- University of Utah (Training Materials and Services)

Join Us!

Support Service Providers (SSP)
Vendor Theater Schedule
Located in Hall B

Come meet caBIG® Support Service Providers (SSP) at the Annual Meeting! The SSP Vendor Theater at the caBIG® World's Fair provides the opportunity to learn about SSP offerings and meet SSP representatives at a scheduled time outside the formal session schedule. The schedule below list times when SSP organizations will be presenting at the Vendor Theater – stop by to learn about the services and products these support providers bring to the caBIG® community.

SSP Vendor Theater Presentation Schedule

SSP	Presentation Time	SSP's Licensed Support Categories
5AM Solutions	Monday, Sept 13 5:30-6:00 PM	• Deployment Support for caBIG® Software Applications • Adaptation and Enhancement of caBIG® -Compatible Software Applications
LabAnswer	Monday, Sept 13 6:15-6:45 PM	• Deployment Support for caBIG® Software Applications • Help Desk Support • Documentation and Training Materials and Services
HeathCare IT, Inc. (HCIT)	Tuesday, Sept 14 12:30-1:00 PM	• Deployment Support for caBIG® Software Applications
SAIC	Tuesday, Sept 14 1:00-1:30 PM	• Adaptation and Enhancement of caBIG® -Compatible Software Applications
INFOTECHSoft, Inc.	Tuesday, Sept 14 5:30-6:00 PM	• Adaptation and Enhancement of caBIG® -Compatible Software Applications
ESAC, Inc	Tuesday, Sept 14 6:15-6:45 PM	• Adaptation and Enhancement of caBIG® -Compatible Software Applications

4d. Tool Demonstrations

World's Fair Technology Demonstrations – Hall C

(All tools are open-source, many are supported by a Knowledge Center, and unless otherwise noted can be found online at https://cabig.nci.nih.gov/tools. More information about support resources can be found in Tab 4c: ESN.)

Biospecimen and Pathology Data Management

Supported by Tissue/Biospecimen Banking and Technology Tools Knowledge Center

- **caTissue Suite** – a web-based biospecimen inventory and tracking tool that may be used by biospecimen resource facilities, regardless of the nature of biospecimen transactions that occur or the type of biospecimens involved in the transaction. Suite also supports integration with associated free-text surgical pathology reports (SPR) and discrete clinical and pathology annotations.
  Presenter: the caBIG® Tissue/Biospecimen Banking and Technology Tools Knowledge Center

- **Laboratory Digital Data Exchange (LIDDEx)** – represents a grass-roots effort driven from the LIS industry itself to address the long-unmet need of clinical laboratories seeking the ability to exchange, electronically, patient results. In the present parlance of electronic interoperability of identified results data, such exchange cannot take place without the a priori expenditure of effort to custom implement a point-to-point interface between the respective originating and receiving institutions. With this model, there is a clear deficit in functionality for exchange of clinical lab data between two "unintroduced" institutions. By use of a formally constrained semantic interoperability model, with anticipated compatibility with caGrid, LIDDEx addresses this deficit with an easily implementable and extensible model, along with model development toolkit, that is readily available for expedited vendor implementation. With the creation of the consortial federation of participating vendors and vendor sites, true laboratory results interoperability becomes possible, agnostic of vendor platform or need for proprietary software interfaces. *For more information please refer to the presenter and visit http://code.google.com/p/liddex*.
  Presenter: Ulysses G.J. Balis, M.D., Ph.D.

Clinical Data and Trials Management

Supported by the Clinical Trials Management Systems Knowledge Center

- **Cancer Adverse Event Reporting System (caAERS)** – an open source software tool that is used to record and manage adverse event data collected during clinical trials. This tool supports regulatory and protocol compliance for safety reporting. This tool also supports service based integration of data from other clinical trials systems.
  Presenter: Paul J. Baumgartner, M.S.

- **Cancer Center Participant Registry (C3PR)** – a web-based application used for end-to-end registration of patients to clinical trials. It can be used in a stand-alone mode or as part of the Clinical suite of applications. Key features of this application include the ability to capture the consent signed date, eligibility criteria, stratification, randomization, and screening. Registration can be done for all types of studies, including companion studies.

The application also includes the ability to handle multi-site clinical trials, integration with NCI Enterprise Services, and Summary 3 reporting.
Presenters: Pankaj Agarwal, M.S., Wesley Wiggins, M.S., and Kruttik Aggarwal

- **Patient Study Calendar (PSC)** – a standards-compliant, web-based application that assists with the management of the activities of subjects on clinical trials. PSC provides the ability to create and edit a standard template to represent the activities defined by a study protocol, use this template to generate and view prospective calendars of subject activities, track the state of activities as a subject progresses through the study, and manage subject calendars as they change during a study. PSC also provides interfaces for managing access to data across a multi-site environment and balancing the workload of Subject Coordinators.
Presenters: Rhett Sutphin and John Dzak

- **Lab Viewer** – a Web based application that is used to view laboratory data in transit between clinical source system (typically in-house clinical chemistry laboratory systems) and clinical trials systems (typically Clinical Data Management System, or CDMS) where such transfer are established using the caBIG® Integration Hub via its Cancer Center Hub Client. Users may browse for laboratory activity by study, participant ID and data range. When Lab Viewer is used in conjunction with the caBIG® Clinical Trials Suite, user can select lab values that are specified in the clinical protocol as labs that must be collected and send them to the CDMS through the caBIG® Clinical Connector to automatically populate electronic Case Report Forms (eCRFs). Additionally, users can select out-of-range lab values to send an alert message to the Adverse Event Reporting systems (caAERS) thereby indicating a potential adverse event
Lab Viewer 2.1 has been enhanced to leverage the National Cancer Institute's services-based enterprise architecture. When users select study personnel and participating organizations in Lab Viewer they are selecting from the curated global list. Through this process, errors and inconsistencies are eliminated and standards are enforced.
Presenter: Naveen Amiruddin

- **Clinical Trials Suite** – a comprehensive set of software tools that facilitates electronic management of clinical trials and associated data, and facilitates comprehensive sharing and integration of clinical research information. Supported clinical trials activities include registering and tracking patients, managing patient activities and calendars, reporting and tracking adverse events, reviewing laboratory data, capturing and cleaning the clinical data, transferring clinical data between applications, and analyzing and reporting on collected data.
Presenter: Bill Dyer

- **Cancer Central Clinical Database (C3D)** – a clinical trials data management system that collects clinical trial data using standard case report form (CRFs) based on common data elements (CDEs). C3D utilizes security procedures to protect patient confidentiality and maintain an audit trail as required by FDA regulations. The application is designed to automate the process of conducting clinical studies on cancer.
Presenter: Joseph Uhimov, Ph.D.

- **caCure** – in collaboration with the NCI, HealthCare IT, Inc., (HCIT) has developed a novel way for researchers to dynamically build their own questionnaires and health surveys. This innovative tool requires little to no developer time and allows researchers to quickly and efficiently build their surveys. The main goals of caCure are to: reduce the development

effort and shorten the time required to add/modify a questionnaire in the user's cohort study website; provide for dynamic creation and extension of a cohort study within a customer styled, tabbed web application; and provide the ability for the customer to create questions that are bound to a structured data model containing pre-existing and/or new data elements. *Early releases will be available from HCIT upon request and the final release will be posted to the caBIG® site.*
Presenters: Suleman Choudhry and Tola Awofolu

- **Form Design Tool Prototype (based on HL7 Reference Information Model (RIM))** – will demonstrate a form design tool prototype that allows a user to design a form template based on HL7 RIM, to use the form template to collect clinical data, and to output the form instance in various form standards including the HL7 CDA (Clinical Document Architecture). *Please refer to the presenters for availability of this tool.*
Presenters: Raghu Chintalapati, M.S., M.B.A. and Cecil Lynch, M.D., M.S.

- **OpenClinica** – an Electronic Data Capture (EDC) and Clinical Data Management (CDM) are trusted by hundreds of industry sponsors, CROs, academic, and government research organizations worldwide, OpenClinica presents a compelling alternative to proprietary, closed systems. *OpenClinica is provided and supported by Akaza Research, LLC. More information can be found at www.openclinica.com.*
Presenters: Cal Collins and Stan Wysocki

Data Sharing and Intellectual Capital

Supported by the Data Sharing and Intellectual Capital Knowledge Center

- **Electronic Data Sharing and Security Framework (EDSSF)** – a web-based tool designed to assist in assessing the sensitivity of a data set and illustrate the legal and regulatory restrictions that may impede dissemination. Constructed around a highly customizable decision tree architecture, the EDSSF's content is readily modifiable and expandable to allow users to tailor it to reflect the needs and concerns of their particular institution or situation.
Presenter: Alex Kanous, M.S.I., J.D.

Imaging

Supported by the Imaging Workspace

- **Annotation Imaging Markup (AIM) Toolkit (Featuring AIM Model and AIM Template)** – the AIM information model provides a supporting infrastructure for creation and collection of medical image annotation needs. Annotation of findings and objects of interest in large clinical or research data collections are fully supported in the model. The AIM project focuses on annotation and markup of DICOM images. These annotations and image markups are information objects that are linked to (but separate from) the images. The AIM information model is, however, compatible with other image formats beside DICOM. This approach is forward-looking, providing an infrastructure for future extension of base functionality that will enable future caBIG® projects, such as annotation for pathology and genomic data, as well as documentation and tracking of quantitative changes to image features. The toolkit consists of the AIM information model, AIM library, and a validation and transformation tool. The model was uploaded to caDSR. A toolkit was released for public

with BSD-style license. The AIM library is written in C++, using DCMTK and Xerces for DICOM and XML creation and manipulation. The library has two logical components: implementation of the AIM Schema as an object model and definition of transformations that can be performed on the AIM object model. The ANIVATR tool is used for validating AIM annotations and transcoding between AIM XML and AIM DICOM SR. *Please refer to the Imaging workspace for more information.*
Presenter: Pattanasak Mongkolwat, Ph.D.

- **eXtensible Imaging Platform (XIP)** – the eXtensible Imaging Platform (XIP™) consists of a set of tools for rapidly developing medical imaging processing and visualization applications. The XIP Wiki site, accessible from *http://www.openxip.org*, describes the tools in more detail. One of the goals for XIP within the caBIG® Imaging program is to allow cancer researchers to easily create complex data analysis programs (e.g. lesion change detection) targeted at specific investigations. By employing the coming DICOM WG-23 Application Hosting interfaces, XIP Applications can be distributed to a variety of installations, thus providing some consistency in data collection for clinical trials. The XIP tools can also be employed to create other types of applications. *Please refer to the Imaging workspace for more information.*
Presenter: Fred Prior, Ph.D.

- **The Cancer Genome Atlas (TCGA) Radiology Workstation (Featuring Clear Canvas, NBIA and AIM Data Service)** – based on an open source imaging workstation from ClearCanvas Inc., Toronto, Canada. The ClearCanvas workstation was enhanced to enable radiologists to annotate and mark-up the different imaging, or phenotypic, characteristics of glioblastoma multiforme patients' magnetic resonance (MR) images on patients whose data were collected for the TCGA project. The goal is to facilitate the collection of imaging phenotypic data in a high structured, machine readable format so that they can be cross correlated with genomic, proteomic, laboratory, clinical, and demographic information to facilitate research and clinical decision support. The TCGA Radiology Workstation has the ability to access the NBIA and the AIM TCGA Data Service in order to search and retrieve imaging studies and AIM XML documents, respectively. Retrieved imaging studies are imported to the workstation automatically. Retrieved AIM XML documents are converted to AIM DICOM SR objects, and then imported to the workstation automatically. If there is more than one AIM XML document associated with an imaging study, all the documents can be included as a part of the study. In addition, the workstation allows a user to quickly create graphical markups and select a predefined and structured set of characteristics describing the brain tumors using the AIM TCGA glioblastoma template. The workstation also allows a user to quickly find AIM annotations in an imaging study and look at basic features such as anatomic entities, imaging observations and quantitative measurements. *Please refer to the Imaging workspace for more information.*
Presenters: Joseph Jen-Sho Chen, M.D. and Vladimir Kleper

- **The Cancer Genome Atlas (TCGA) Genotype/Phenotype Analytics (Featuring the caIntegrator2 TCGA Cancer Cell Study)** – the TCGA Radiology project utilizes multiple CBIIT/caBIG® technologies together to create a practical system to capture diagnostic imaging "knowledge" in a structured, standardized manner and to allow for the integration with genomic and clinical data. The project involved neuro-radiologists reading eighty-eight TCGA Radiology cases from the National Biomedical Imaging Archive (NBIA) that correspond to data from the December, 2009 TCGA Cancer Cell article. The results were published on the caGrid using three independent data services and a set of translational queries were created for the study using caB2B and caIntegrator. This demo will focus on

using caIntegrator to query the combined data set and analyze survival results based on a combination of clinical factors, genomic categorizations, and imaging observations. *Please refer to the Imaging workspace for more information.*
Presenter: Chris Piepenbring

- **Middleware/ Virtual Picture Archiving and Communication Systems (PACS)** – the In Vivo Imaging Middleware (better known as IVI middleware) is a set of open source tools and middleware, which provide interoperability between DICOM and the Grid. IVIM development is driven by common use cases from radiology and pathology for common activities such as central review in image based clinical trials, data and application sharing in secure high performance environments. It is designed to provide grid-based, federated access to existing DICOM-based data repositories and analytical resources and to facilitate development of grid-enabled in vivo imaging applications in caBIG®. It is layered over the grid infrastructure provided by caGrid, leveraging caGrid's core services, toolkits, and wizards for the development and deployment of community-provided services and APIs for building client applications. It includes three new notable components: RESTful source code generator tool (RSGT); Annotation and Image Markup service at Emory; Web Application for DICOM Data Service. *Please refer to the Imaging workspace for more information.*
Presenters: Ashish Sharma, Ph.D. and Tony Pan, M.S.

- **National Biomedical Imaging Archive (NBIA)** – provides the cancer research community, industry, and academia with public access to DICOM images, image markup, annotations, and rich meta data. NBIA provides Web-based access to de-identified DICOM images, markups, and annotations using role-based security. *Please refer to the Imaging workspace for more information.*
Presenter: Brian Hughes

- **Semantically-aware Image Annotation (iPAD) using AIM** – an annotation tool for radiology images, to make the semantic content (the meaning and other key metadata) explicit and machine-accessible. iPAD, a plug-in to the popular OsriX image viewing workstation, implements the AIM standard (Annotation and Image Markup) of the caBIG® project. The problem iPAD addresses is the medically-important content in images – the anatomy, radiology findings, and quantitative features, are not recorded in a way that machines can access the information – they are either in graphical overlays or in unconstrained text in the DICOM header. With iPAD, we can store and share information about image anatomy, visual features, and quantitative assessments in a computable format. The user draws an ROI on the image, and then types into the iPAD, recording the observations and anatomic location information. Behind the scenes, iPAD records all this information in the AIM format (XML). This can then be stored in a database or serialized to DICOM-SR. *Visit http://bimm.stanford.edu/main/ipad*
Presenter: Daniel Rubin, M.D., Ph.D.

Infrastructure

Supported by the caGrid Knowledge Center and the Vocabulary Knowledge Center

- **NCI Metathesaurus** – a comprehensive biomedical terminology database that contains 1,400,000 concepts mapped to 3,600,000 terms with 17,000,000 relationships. NCI Metathesaurus contains most terminologies used by NCI for clinical care, translational and

basic research, and public information and administrative activities, including most public do70

- main terminologies from the National Library of Medicine's UMLS Metathesaurus as well as a growing number of other cancer-related and biomedical terminologies.
  Presenters: Robert Freimuth, Ph.D., Jyoti Pathak, Ph.D. and Cui Tao, Ph.D.

- **NCI Thesaurus** – published monthly by NCI, this reference terminology and biomedical ontology is used in a growing number of NCI and other systems. It covers vocabulary for clinical care, translational and basic research, and public information and administrative activities. The NCI Thesaurus provides definitions, synonyms, and other information on nearly 10,000 cancers and related diseases, 8,000 single agents and combination therapies, and a wide range of other topics related to cancer and biomedical research. It is maintained by a multidisciplinary team of editors, who add about 900 new entries each month.
  Presenters: Robert Freimuth, Ph.D., Jyoti Pathak, Ph.D. and Cui Tao, Ph.D.

- **Common Terminology Services II (CTS2)** – specifies a set of service interfaces that standardize the functional requirements of a terminology server to allow the representation, access and maintenance of terminology content. The scope of functionality includes administration, search and query, authoring and maintenance and mapping support. As a joint standards development activity between Health Level Seven (HL7) and the Object Management Group (OMG) as a multi-organizational project, CTS 2 seeks to provide a standardized interface to controlled terminology resources.
  Presenters: Robert Freimuth, Ph.D., Jyoti Pathak, Ph.D. and Cui Tao, Ph.D.

- **NCI Enterprise Vocabulary Services (LexEVS)** – a collection of programmable interfaces based on standards including Common Terminology Services (CTS) Release 1.2 and CTS Release 2, which provide users the ability to access controlled terminologies supplied by the NCI Enterprise Vocabulary Services (EVS) Project. The controlled terminologies hosted by the NCI EVS Project are published via the Open-Source LexEVS Terminology Server. As a caCORE Software Development Kit (SDK) generated system, LexEVS provides a set of tools that can be used by an application developer to create a caCORE-like system.
  Presenters: Robert Freimuth, Ph.D., Jyoti Pathak, Ph.D. and Cui Tao, Ph.D.

- **LexWiki/CSHARE** – a collaborative effort led by Mayo Clinic for development of a collaborative authoring platform for large-scale biomedical terminologies. LexWiki currently is at the core of community-based development of Biomedical Grid Terminology (BiomedGT).The CDISC Shared Health and Research Electronic Library (CSHARE) wiki, an extension of LexWiki, is a collaborative authoring environment that enables community members to harmonize data elements from multiple organizations.
  Presenters: Robert Freimuth, Ph.D., Jyoti Pathak, Ph.D. and Cui Tao, Ph.D.

- **caGrid Portal** – provides a visual display of the services that are running on the caGrid infrastructure and also institutions that are participating in the caBIG® program. From this portal you can find caBIG® services, participants, and points of contact, caBIG® software tools, database technologies and web-based applications. You can also query caGrid data services, share your queries, and search for queries of other community members.
  Presenter: Musharaf Rashid

Molecular Characterization

Supported by the Molecular Analysis Tools Knowledge Center

- **caArray** – a standards-based data management system that supports the annotation and exchange of microarray data using a federated model of local installations whose data are shareable across caGrid. The programmatic interfaces of caArray allow tools, such as geWorkbench, GenePattern, and caIntegrator, to seamlessly access data for analysis. caArray supports the bulk upload of data sets using the MAGE-TAB standard. In addition, caArray supports the export of data in a format suitable for submission to GEO. The security features of the system allow users to control access to their data sets at the sample or experiment level and allow read or read/write access to be given to individuals, collaboration groups, or the public, as appropriate.
  Presenters: Don Swan and Zhong Li, Ph.D.

- **caCorrect (chip artifact CORRECTion**) – a software program that improves the quality of collected microarray data, ultimately leading to improved cancer biomarker selection. Widely used Affymetrix microarrays contain millions of probes (a 25-oligo sequence) used to detect mRNA expression levels. It is easy for regions of probes to become marred by artifacts and noise. These unusable portions are typically the result of experimental variations by different laboratory technicians or errors that create scratches, edge effects and bubble effects on the data. caCORRECT removes the noise and artifacts from the data, while retaining high-quality genes on the array. The software can also effectively recover lost information that has been obscured by artifacts. *Refer to the tool presenter for support information.*
  Presenter: Todd Stokes, Ph.D.

- **caIntegrator2** – a web-based software package that allows researchers to set up custom, caBIG®-compatible web portals to conduct integrative research, without requiring programming experience. These portals bring together heterogeneous clinical, microarray and medical imaging data to enable multidimensional analysis.
  Presenters: Karen A. Ketchum, Ph.D., JP Marple and Shine Jacob, M.S.

- **cancer Bench-to-Beside (caB2B)** – a grid client that provides query templates that allow easy search and retrieval of microarray data (from caArray), imaging data (from the National Biomedical Imaging Archive), specimen data (from caTissue) and nanoparticle data (from caNanoLab) across the grid. *Refer to the Molecular Workspace for support information. Refer to the tool presenter for support information.*
  Presenters: Baris Suzek, M.S. and Jim Humphries

- **Cancer Bioinformatics Infrastructure Objects (caBIO)** – provides Application Programming Interface (API) access to an integrative view of biological annotations originating from a variety of data sources. Information from key annotation providers – including UniGene, Entrez Gene, and UniProt - is stored in the caBIO infrastructure and is updated on a semi-monthly basis. Recent additions to caBIO have included annotations from popular microarray platforms, curated disease-gene-agent information from the Cancer Gene Index and Canada DrugBank, pathway interactions from the Pathway Interaction Database (PID).
  caBIO data is made accessible through a variety of APIs including caGrid API and the RESTful, SOAP, and Java APIs. caBIO is also searchable through a variety of graphical user interfaces including: a caBIO home page which provides a utility (FreestyleLM) for performing "Google®-like" searches; a caBIO Portlet for easily accessing caBIO data via the

caGrid Portal by performing pre-defined templated searches; and a caBIO iPhone App for retrieving molecular annotations via the iPhone, iPod, and iPad devices. The caBIO project is currently engaged in a pilot effort involving the development of a molecular annotation service leveraging the NCI Enterprise Conformance and Compliance Framework (ECCF). *Refer to the tool presenter for support information.*
Presenter: Jim Sun, Ph.D.

- **Cancer Genome Workbench** – provides a variety of tools for visualizing sample-level genomic data. Built originally on the UCSC genome browser, CGWB has extended the UCSC platform to include tracks for individual samples, as well as heatmaps, scatter plots, sequence traces, next-generation sequence alignments, and 3D protein structures. CGWB supports a number of high-throughput genomic data types, including copy number alterations, genotypes, somatic mutations, mRNA and miRNA expression, and methylation. The heatmap viewer offers interactive displays of gene expression and copy number changes along with clinical features. The next-generation sequence viewer, Bambino, allows users to examine the sequence alignment quality at base level and to identify SNPs/indels from next-generation sequence mapping files in SAM/BAM format. Data in CGWB is organized in projects and sub-projects. *CGWB is currently hosting data from the following projects: The Cancer Genome Atlas (TCGA), Therapeutically Applicable Research to Generate Effective Treatments (TARGET), the Sanger Institute's Catalog of Somatic Mutations in Cancer (COSMIC), the NCI60 cell lines, and the GlaxoSmithKline cell lines.* Much of the data in CGWB is available on caGrid. With comprehensive genomic alteration data from large numbers of tumor samples and cell lines, CGWB will help researchers gain new insight into cancer biology.
Visit http://cgwb.nci.nih.gov and refer to the tool presenter for support information.
Presenter: Chunhua Yan, Ph.D.

- **Cancer Models Database (caMOD)** – provides information about animal models for human cancer to the public research community. caMOD allows users to retrieve the making of models, their genetic descriptions, histopathology, images, microarray data, and therapeutic trials in which the models were used. *Visit http://cancermodels.nci.nih.gov* and *refer to the tool presenter for support information.*
Presenter: Sima Pandya, M.S.

- **Cancer Nanotechnology Laboratory Portal (caNanoLab)** – a portal for the annotation of nanoparticles with characterizations resulting from physical and in vitro nanoparticle assays and the sharing of these characterizations and associated nanotechnology protocols in a secure fashion. *Refer to the tool presenter for support information.*
Presenters: Sharon Gaheen, M.B.A. and Sue Pan

- **geWorkbench** – provides a platform for genomic data integration. It brings together analysis and visualization tools for gene expression, sequence, protein structure, pathway, and gene annotation data. It gives scientists transparent access to a number of external data sources and algorithmic services, and integrates them with its own wide selection of built-in tools, allowing easy flow of data between all.
Presenter: Kenneth Smith, Ph.D.

- **GenePattern** – freely available platform for integrative genomics, it provides access to a broad array of computational methods used to analyze genomic data including over 100 computational and visualization tools for the analysis of gene expression, SNP and

proteomic data. Its extendable architecture makes it easy for computational biologists to add analysis and visualization modules, which ensures that GenePattern users have access to new computational methods on a regular basis. GenePattern includes an integrated workflow engine to facilitate the creation and sharing of reproducible in-silico methodologies.
Presenter: Jared Nedzel, M.S.

- **Pathway Interaction Database (PID)** –The National Cancer Institute (NCI) in collaboration with Nature Publishing Group has established the Pathway Interaction Database (PID) in order to provide a highly structured, curated collection of information about known biomolecular interactions and key cellular processes assembled into signaling pathways. PID is aimed at the cancer research community and others interested in cellular pathways, such as neuroscientists, developmental biologists, and immunologists. The database focuses on the biomolecular interactions that are known or believed to take place in human cells. It can be browsed as an online encyclopedia, used to run computational analyses, or employed in ways that combine these two approaches. In addition to PID's predefined pathways, search results are displayed as dynamically constructed interaction networks. These features of PID render it a useful tool for both biologists and bioinformaticians. Visit *http://pid.nci.nih.gov and refer to the tool presenter for support information.*
Presenter: Jeffrey Buchoff

- **Taverna Workbench** – allows users to construct complex workflows. The workflows can consist of multiple types of components, each type of component is called a processor. These components which may be located on different machines. Their execution is orchestrated by Taverna Engine (run from the workbench), and the results are gathered and shown in the workbench. The current versions of Taverna support many types of processors, including apiconsumer (for running Java methods), beanshell processor, biomart processor, biomoby processor, R script processor, soaplab processor, WSDL processor. *Refer to the tool presenter for support information.*
Presenter: Wei Tan, Ph.D.

- **The Cancer Genome Atlas (TCGA) Data Portal** – a comprehensive and coordinated effort by the National Cancer Institute and the National Human Genome Research Institute to accelerate our understanding of the genetics of cancer. In 2009, the NIH announced the expansion of the scope of the TCGA Research Network to include more than 20 tumor types over the next five years. Each cancer will undergo comprehensive genomic characterization that incorporates powerful bioinformatics and data analysis components, that will enable researchers to further mine the data generated by TCGA to improve prevention, diagnosis, and treatment of cancer. The center of the informatics efforts is the TCGA Data Coordinating Center (DCC), where the information generated by the Biospecimen Core Resources, the Genome Sequencing Centers, the Cancer Genome Characterization Centers, and the Genome Data Analysis Centers is centrally managed and entered into public databases. The TCGA Data Portal allows access to the DCC data. Most data within the Data Portal is publicly accessible; however, access to some clinical data requires user certification. Visit *http://cancergenome.nih.gov and refer to the tool presenter for support information.*
Presenter: John M. Greene, Ph.D.

Population Science

Supported by the Population Sciences SIG

- **Grid Enabled Measures (GEM)** – a dynamic web-based portal built upon the caBIG® platform. It provides a forum for a virtual community of researchers to interact with each other using web 2.0 capabilities. The goals of GEM are to promote the use of standardized measures-- which are tied to theoretically-based constructs --and facilitate the ability to share harmonized data resulting from the use of standardized measures. Through the use of these measures and common elements, prospective meta-analyses will be possible. *Refer to the tool presenter for support information.*
  Presenter: Rick Moser, Ph.D.
- **Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE)** – the purpose of the PRO-CTCAE project is to develop an electronic-based system for patient self-reporting of symptom adverse events (AEs) listed in the CTCAE in an effort to improve the accuracy and precision of grading of this class of AEs. The accurate reporting of AEs that occur to patients on clinical trials is a federal requirement that facilitates evaluation of new therapies. *Refer to the tool presenter for support information.*
  Presenter: Kate Castro, R.N., M.S. and Mehul Gulati

- **Population Science Grid 1.0 (PopSciGrid 1.0)** – based on a conceptual framework for cyber-enabled collection, harmonization, and analysis of population health data, NCI is developing the PopSciGrid consumer health information portal. The PopSciGrid Portal will demonstrate how tobacco prevalence and policy data can be integrated, visualized, and communicated to help empower communities and decision-makers. GEM and the PopSciGrid Portal, both of which are integrated with NCI's cancer Biomedical Informatics Grid (caBIG®), focus on supporting greater transparency, scientific collaboration, and community participation in the cancer prevention and control. Grid-enabled applications such as these can target specific public health stakeholders for cancer prevention and control. *Refer to the tool presenter for support information.*
  Presenter: Abdul Shaikh, Ph.D.

5. Speaker Biographies

Speaker Biographies

R. Mark Adams, Ph.D.

adams_mark@bah.com

R. Mark Adams, Ph.D., has been in the field of bioinformatics for more than 15 years with extensive experience in the design and implementation of large-scale informatics systems, including regulated systems for use in clinical trials and the management of high-throughput laboratory processes. Additionally, his theoretical work has encompassed abstract mathematical and computational analysis of molecular genetic and protein structure. His scientific work has resulted in many patent applications, with several patents granted. As a biotechnology executive, Dr. Adams has extensive experience in the strategic leadership and management of scientific and technical teams, having managed teams of up to 50 biologists and computer scientists designing and implementing high-throughput molecular biology and bioinformatics systems. In his role as vice-president of a public biotechnology company, Dr. Adams served on the executive committee, chartered with creating and executing a strategic plan for cancer pharmacogenomics. He is currently the scientific program manager for the National Cancer Institute cancer Biomedical Informatics Grid program (caBIG®) and the leader of the Booz Allen Hamilton Biomedical Informatics group. Dr. Adams is an active member of the field of bioinformatics where he speaks and publishes frequently. He is the past editor of the *Journal of Applied Analytics in Consulting*, and *Briefings in Bioinformatic*s.

Beverly Albury

alburyb@sjchs.org

Beverly Albury, NCCCP IT Lead, has 20 years of healthcare experience, 12 years in information technology in providing direct support and expert assistance to customers in the implementation and use of healthcare information technology applications. Beverly Albury, has been engaged in NCCCP information technology consistently since the selection of St. Joseph's/Candler as one of the NCCCP sites. She was instrumental in assisting to define the requirements and use cases for an oncology extended EHR which lead to the development of Clinical Oncology Requirements for the EHR (CORE) document. Ms. Albury is currently one of the chairs of the NCCCP Recommendation White Paper Committee where she provides guidance for the implementation of Information Technology products at NCCCP sites, including but not exclusive to caBIG® products.

Mariano Alvarez, Ph.D.

malvarez@c2b2.columbia.edu

Mariano Alvarez, Ph.D., studied biology and got his Ph.D. in chemistry in the University of Buenos Aires, Argentina, by investigating the interaction between human melanoma cells and the extracellular matrix. He joined Dr. Califano's Lab at Columbia University in 2006, where he initially did postdoctoral training for 2 years. He is working now in the inter-phase between biology and computer science trying to identify driver genes (or Master Regulators) of tumor development and progression and drug mechanisms of action by using systems biology approaches.

Christo Andonyadis, D.Sc.

andonyac@mail.nih.gov

Christo Andonyadis, D.Sc., has over 25 years' experience in systems development and information science, and has focused specifically in the clinical research domain for the last 10 years. As Chief Architect for Clinical Science, he is responsible for the design, development, and support of clinical trials applications and services in the NCI CBIIT's portfolio – supporting the intramural research community as well as the caBIG® community. Dr. Andonyadis chairs the Clinical Science Composite Architecture Team (CSCAT), which has been established to govern the clinical research domain. The CSCAT has representatives for each of the five dimensions of the Reference Model for Open Distributed Processing (RM-ODP, ISO/IEC IS 10746|ITU-T X.900) – Enterprise, Information, Computational, Engineering and Technical. Dr. Andonyadis has been actively involved with the caBIG® Clinical Trials Management Systems (CTMS) Workspace since its inception – having served as NCI CBIIT lead to the CTMS workspace as well as lead to the Structured Protocol Special Interest Group (SIG) and the Regulatory Reporting SIG. He provides guidance and direction for various SIG activities including white paper authoring, functional requirement gathering, strategic planning, and goal definition.

Michael E. Berens, Ph.D.

Michael E. Berens, Ph.D., is Associate Director for the Van Andel Research Institute, as well as Director of the Cancer and Cell Biology Division at the Translational Genomics Research Institute, TGen (where he is Senior Investigator in the Brain Tumor Unit). His long-standing professional interest is in cancer invasion, which leads to the ultimate cause of cancer-related death: metastasis. While working at Barrow Neurological Institute, he developed & patented a research-based technology for innovative studies of cell migration; that technology was a catalyst which led to the founding of Avolix in 2000. He is cofounder and currently serves as the chair of the Scientific Advisory Board of Avolix Pharmaceuticals, Inc. In 2002, Dr. Berens served as a member of Governor Jane Dee Hull's (Arizona) task force for bioscience; in that capacity he assisted in garnering support in Arizona to launch the Translational Genomics Research Institute (TGen): approximately $100M public-private contributions, strategic alliances with Arizona's three public universities, engagement with national thought leaders on the founding board of directors, and recruitment of entrepreneurial translational researchers. In 2003, Dr. Berens was founding CEO of the International Genomics Consortium, Phoenix, securing $16M from public-private (pharmaceutical) sources to build public genomic databases of human cancers. Following the award of his Ph.D. from the University of Arizona (1982), Dr. Berens' academic career has included appointments at the University of Zurich, the Bowman Gray School of Medicine of Wake Forest University, the University of California at San Francisco, and the Barrow Neurological Institute, where he worked 12 years as Senior Investigator and Director of Neurology Research. His translational research program on brain tumors includes preclinical therapy development, novel treatment target discovery, and basic science of malignant cell motility. His research program includes international collaborations, and is funded by the National Institutes of Health and private medical Foundations. Dr. Berens holds four patents. He serves on boards for scientific journals, governmental agencies, professional societies, and non-profit organizations. He is active in the technology and public policy sectors, and is a past Chairman of the Arizona Technology Council. Mike has been married to Patty for 33 years; they have two children.

Daniel Brat, M.D., Ph.D.

dbrat@emory.edu

Daniel Brat, M.D., Ph.D., is a Professor and Vice Chair of Translational Programs in the Department of Pathology and Laboratory Medicine at Emory University School of Medicine. He received his MD and PhD degrees from the Mayo Medical and Graduate Schools and then completed a Residency in Anatomic Pathology and Fellowship in Neuropathology at the Johns Hopkins Hospital. Dr. Brat is a diagnostic neuropathologist for the hospitals of both Emory Healthcare and Children's Healthcare of Atlanta. He is also a member of the Winship Cancer Institute, where his lab investigates mechanisms of brain tumor progression. He has led the neuropathology investigations of the TCGA for glioblastoma and currently chairs the TCGA efforts for lower grade gliomas. These projects interface with the Emory In Silico Center for Brain Tumor Research, for which Dr. Brat serves as the scientific PI. Dr. Brat also directs the Research Pathology Lab at Emory, which has expertise in standard and advanced histologic techniques, including whole slide scanning for digital pathology, tissue micro-arrays and laser capture microscopy. Dr. Brat serves on numerous committees that oversee diagnostic practice and translational investigation in oncology, including the College of American Pathologists (CAP) Neuropathology Committee, the WHO Classification Committee for Brain Tumors, the Radiation Therapy Oncology Group (RTOG) and Children's Oncology Group (COG). He sits on 7 journal editorial boards in pathology and oncology and reviews grant proposals for the NIH. His scholarly activity has resulted in over 120 publications, the Lucien J. Rubinstein and the Matthew T. Moore Awards from the American Association of Neuropathologists and the Benjamin Castleman Award from the United States and Canadian Association of Pathologists.

Terry Braun, Ph.D.

tabraun@eng.uiowa.edu

Terry Braun, Ph.D., received both his B.S. degree and his M.S. degree in Electrical and Computer Engineering from the University of Iowa in Iowa City. He then obtained his Ph.D. in Genetics from the University of Iowa. Currently, he is an Associate Professor in the departments of Ophthalmology and Visual Sciences, and Biomedical Engineering. With a background in high-performance computer architecture, parallel/distributed computing systems, algorithms, and software engineering, he now specializes in computational biology, genomics, bioinformatics, and genetics. He is director of the Coordinated Laboratory for Computational Genomics (CLCG), and his laboratory is developing computational tools and novel techniques for disease gene identification, prioritization, and annotation for disorders that include age-related macular degeneration, glaucoma, retinitis-pigmentosa, autism, Bardet-Biedl Syndrome, hypertension and neuroendocrine tumors.

Lawrence Brem, M.S.

breml@mail.nih.gov

Lawrence Brem, M.S., is the Technical Program Manager for both the caBIG® Architecture and Vocabulary Workspaces. In this role he oversees the development of the NCI Funded caGRID projects including the platform and tooling, the caGRID Workflow and the caGRID Portal and the Semantic Projects including LexEVS, caDSR and Curation efforts. Prior to joining SAIC-F in 2008, Mr. Brem developed and oversaw computer middleware solutions for the DoD Testing and Training Community. Mr. Brem received his Bachelors in Electrical Engineering from Worcester Polytechnic Institute and his Masters in Electrical Engineering from the University of Virginia.

Elaine L. Brock, M.H.S.A., J.D.

megmorin@umich.edu

Elaine L. Brock, M.H.S.A., J.D., is the Senior Associate Director of the Division of Research Development and Administration (DRDA) and, for some years, was also the Director of the Medical School Office of Technology Transfer and Corporate Research (TTCR) at the University of Michigan. Within DRDA, Brock is responsible for the University-wide pre-award office including oversight of all aspects of sponsored projects administration such topics as clinical research, conflict of interest, intellectual property, export control, and data rights. She has served on statewide, regional, and national committees developing policies and procedures affecting university-industry interactions. She is a frequent presenter at national conferences on conflict of interest and various other aspects of industry-university relationships. Brock has been an active participant in the Data Sharing and Intellectual Capital Workspace of the Cancer Biomedical Informatics Grid (caBIG®). As Director of the DSIC Knowledge Center, Brock is expanding discussion about issues related to data sharing and developing solutions and approaches to identified barriers. Brock is currently a member of the Contracts and Intellectual Property Committee and the Board of Directors of the Council on Governmental Relations (COGR). She is also an active participant in the University Industry Demonstration Partnership. Brock has a B.A. from the State University of New York at Buffalo with a double major in psychology and political science, an M.H.S.A. from the University of Michigan in Health Planning and Administration, and a J.D. from Michigan State University College of Law. She is licensed to practice law in the State of Michigan.

Bartley Brown, Ph.D.

bbrown@eng.uiowa.edu

Bartley Brown, Ph.D., has been the caBIG® Center Deployment Lead at the University of Iowa since 2007. He is also a software developer for the University of Iowa's Center for Bioinformatics and Computational Biology (CBCB) and has worked in this capacity since 2000. Bart has been heavily involved with the deployment and adaptation of caTissue Suite for the University's Tissue Procurement Core and has provided consultation to other University entities regarding caTissue Suite and caBIG® tools in general. He and Dr. Thomas Casavant, director of the CBCB, have participated in the Documentation and Training Workgroup for approximately two years. Dr. Brown's involvement with caBIG® started in 2004 when he assisted in the process of creating a silver-compatible version of TrAPSS, a tool developed by the CBCB. TrAPSS stands for **Tr**anscript **A**nnotation **P**rioritization and **S**creening **S**ystem and is a system comprising several tools that aid in the search for the genetic mutations that are linked to expression of a disease phenotype.

Jennifer Brush, M.S.

jbrush@scenpro.com

Jennifer Brush, M.S., is a Project Manager with ScenPro with over 16 years of experience in project management and system requirements analysis. She is an expert in requirements elicitation, analysis and management; training and documentation; and consensus building among clients, colleagues and stakeholders. In addition to project management, Ms. Brush trains, motivates and mentors both analysts and trainers and has deep domain knowledge in the areas of ISO/IEC 11179 Standard-based metadata

and proficiency in developing and providing effective training for adult learners. Ms. Brush has worked with NCI CBIIT since 2003 and has led ScenPro's training development activities with NCI CBIIT since 2005.

Miguel Buddle, M.S.

buddle_miguel@bah.com

Miguel Buddle, M.S., is a Lead Associate with Booz Allen Hamilton and is the Manager of the caBIG® Program Management Office. In addition to his program management duties, Mr. Buddle also serves on the Tiger Team for the Enterprise Support Network and is the lead for the caBIG® Deployment activity. After having spent many years in Pathology Informatics, Mr. Buddle came to Booz Allen in 2007 from the United States Military Cancer Institute (USMCI), where he served as the Information Systems Manager and was responsible for managing USMCI's efforts to harmonize cancer research IT infrastructures across the military services.

Kenneth H. Buetow, Ph.D.

kenneth.buetow@mail.nih.gov

Kenneth H. Buetow, Ph.D., has a multi-disciplinary scientific career that has focused for more than 20 years on understanding the role of genetics in complex human diseases such as cancer, and on applying sophisticated informatics technologies to solve major biomedical challenges. In his current role of National Cancer Institute Associate Director responsible for Bioinformatics and Information Technology, he initiated and oversees the caBIG® (cancer Biomedical Informatics Grid) program, a groundbreaking initiative built to connect the entire cancer community in a "World Wide Web" of biomedical research. caBIG® has pioneered the infrastructure and a portfolio of tools that enable organizations and individual researchers to securely share biomedical data, and its capabilities serve as a demonstration of the connectivity required for Personalized Medicine. Dr. Buetow also serves as the Director of the NCI Center for Bioinformatics and Information Technology (NCI CBIIT), which is responsible for maximizing the interoperability and integration of NCI research. He is also the Chief of the Laboratory of Population Genetics (LPG), where his group applies genomics to increase our understanding of the genetics of complex phenotypes. In addition to serving on the governing and advisory boards for numerous government organizations, academic institutions, and scientific and medical societies, Dr. Buetow has published more than 160 scientific papers. His recent honors and awards include The Editor's Choice Award from Bio-IT World (2008), The Federal 100 Award (2005), The NIH Award of Merit (2004) and the NCI Director's Gold Star Award (2004). Dr. Buetow received a B.A. in biology from Indiana University in 1980 and a Ph.D. in human genetics from the University of Pittsburgh in 1985.

Braulio J. Cabral

braulio.cabral@nih.gov

Braulio J. Cabral is the NCI-CBIIT Enterprise Security Program Coordinator. In his role Mr. Cabral coordinates activities related to security compliance with federal mandates such as FISMA, OMB, HIPAA, etc., activities related to integration of security into the Software Development Lifecycle, coordination of systems certification and accreditation activities, as well as security awareness activities across caBIG® community. Prior to joining the Enterprise Security Program, Mr. Cabral served in several IT positions at GXS, Inc. (formerly GE Information Services) a B2B consulting and service provider firm located in Gaithersburg, MD. During his 15 years at GXS, Inc. he served as manager for

customer services technical support, lead systems integration and implementation, technical consultant for integration and information security, manager second tier global service center where his team provided 24/7 engineering support for Asia Pacific, Europe, Africa and the Americas. Mr. Cabral holds a Bachelor of Science in Computer and Information Systems from Strayer University in Washington DC; a Master of Science in Information Technologies with a concentration in Object Oriented programming, from Regis University in Colorado, and a Master of Science in Management Information Systems Security and Project Management from Colorado Technical University; he is a Ph.D. in Applied Management student at Walden University, and is a Sherwood Applied Business Security Architecture (SABSA) certified security architect. He is a member of the Information System Security Association (ISSA), the Information Systems Audit and Control Association and the Project Management Institute (PMI).

Andrea Califano, Ph.D.

Andrea Califano, Ph.D., pursues interests that are broadly defined within the field of Systems Biology, with specific application to human malignancies. In particular his lab has spearheaded early efforts to assemble genome-wide, context-specific maps of molecular interactions in human cells, by integrating several reverse engineering approaches. These maps have shown significant promise in the rational elucidation of both physiological and pathological phenotypes. In 2003, Dr. Califano joined Columbia University as Professor of Biomedical Informatics, with appointments in the Department of Biomedical Informatics and in the Institute for Cancer Genetics. Over the last few years, his lab has assembled biochemically validated, genome-wide map of transcriptional and post-transcriptional interaction in several human cell contexts, including B cell, Breast Carcinoma, Glioma, and normal and tumor-related Stem Cells. These maps are being extensively used for the unbiased dissection of dysregulated pathways in related human malignancies. The Califano lab integrates the development of analytical methodologies with high-throughput experimental assays necessary for data generation and biochemical/biological validation.

Regina Cer, M.S.

cerr@mail.nih.gov

Regina Cer, M.S., is a bioinformatics analyst and has been with the Advanced Biomedical Computing Center (ABCC) in the Bioinformatics Support Group for three years. Her work involves solving biological problems using databases and visualizing the solution through web-based technologies. After joining the In Silico Research Center of Excellence project, she has been the main person for maintaining the non-B data work. She is also serving as the ABCC's point of contact for the ISRCE project.

Joseph Jen-Sho Chen, M.D.

josephjschen@gmail.com

Joseph Jen-Sho Chen, M.D., is a board certified radiologist and currently serves as the life science workspace subject matter expert (SME). Dr. Chen has extensive experience and expertise in imaging informatics, and has responsibility for helping the community formulate life science based enterprise use cases and leveraging cancer Biomedical Informatics Grid (caBIG®) tools. In addition, he works in the Department of Diagnostic Radiology and Nuclear Medicine at the University of Maryland School of Medicine, as a visiting instructor in cardiothoracic and MR imaging. Dr. Chen has presented 22 national presentations and posters, authored or co-authored 16 peer-reviewed scientific

papers, and 53 book chapters. His areas of expertise include rapid assessment of chest pain in the Emergency Department using computed tomography (CT), computer-aided detection (CAD) and evaluation of lung cancer, and quantitative analysis of CT imaging. Dr. Chen has trained under Dr. Eliot L. Siegel, the Vice-Chairman of Imaging Informatics at the Department of Diagnostic Radiology and Nuclear Medicine at the University of Maryland School of Medicine, Chief of Imaging Services at the Department of Veterans Affairs Maryland Health Care System, and lead SME of the Imaging Workspace at caBIG® in the National Cancer Institute's (NCI) Center for Biomedical Informatics and Information Technology (CBIIT).

Raghu Chintalapati, M.S., M.B.A.

rchintal@ekagrasoft.com

Raghu Chintalapati, M.S., M.B.A., is the Chief Technology Officer at Ekagra Software Technologies. Mr. Chintalapati is currently a Lead Architect at caBIG® working on the Semantic Infrastructure. Prior to his current role, Mr. Chintalapati was the project manager of both caDSR and EVS. Mr. Chintalapati also held the position of Director at Oracle Corporation in the solutions architecture group. Mr. Chintalapati received an M.S., from Wayne State University, and his MBA from Cornell University. Mr. Chintalapati has been actively involved in caBIG® since 2008, and has led the caGrid deployment team and the caXchange teams.

Laurence P. Clarke, Ph.D.

lclarke@mail.nih.gov

Laurence P. Clarke, Ph.D., has been the Branch Chief for Imaging Technology Development for the Cancer Imaging Program, Division of Cancer Treatment and Diagnosis, NCI, since 1999. In this capacity, Dr. Clarke is responsible for the development of initiatives for new and emerging imaging platforms and methodologies as applied to cancer, with a particular emphasis on the role of imaging as a biomarker, and the importance of physical imaging standards. Dr Clarke is also on a detail assignment at NIBIB and is a guest scientist at NIST. His responsibilities at NCI, NIBIB and NIST include initiatives that support public research resources for assessing new biomarker imaging methods including the evaluation of quantitative imaging methods for the measurement of drug response in particular. Before joining NCI, Dr. Clarke was a Professor of Radiology, with adjunct appointments in computer science and physics at the University of South Florida (USF), and was a full member and imaging program leader at the H. Lee Moffitt Cancer and Research Center at USF. He is a Fellow of the ISMRM and AAPM. He graduated with a Ph.D. in physics at the National University of Ireland (1978) and a M.Sc. degree from Queens University of Belfast, Ireland (1968).

Michael A. Collins, M.S.

michael.collins@fccc.edu

Michael A. Collins, M.S. is Director of Research Informatics at the Fox Chase Cancer Center (FCCC). Since 2004 he has lead the design and development of software architecture supporting informatics tools for translational research. His research interests are using data integration and service-oriented architecture to accelerate scientific discovery and improve data quality. He is a funded participant as a Subject Matter Expert for the caBIG® Population Sciences Special Interest Group and Architecture Workspace and serves as the caBIG® Deployment Lead for FCCC.

Deborah Collyar

Deborah@tumortime.com

Deborah Collyar has been a leader in cancer patient advocacy since 1991, utilizing successful business and computer skills to bridge research gaps between science and patients. Her patient advocacy work encompasses many diseases and policies at grassroots, regional and national levels, as well as guidance to international organizations and agencies. Ms. Collyar works in depth with collaborators from many government agencies, academic and private institutions, companies, professional societies, non-profits, advocacy organizations, and cancer patients. Ms. Collyar founded the PAIR international network; started patient advocacy in CALGB, and serves as Chair of their Committee on Advocacy, Research Communication, Ethics, and Disparities (CARE). She has served as Chair of the Patient Advisory Board to the Coalition of Cancer Cooperative Groups and as Program Director for the SPORE Patient Advocate Research Team (PART) Program, expanding the UCSF pilot she helped to create. Ms. Collyar currently participates in NCI's caBIG® Data Sharing and Intellectual Capital workspace, Investigational Drug Steering Committee (IDSC), Cancer Human Biobank (caHUB), and NCI Experimental Therapeutics (NExT). She has also served on many NCI/NIH committees (e.g. Board of Scientific Counselors); in AACR and ASCO workshops and committees; on ethical and executive advisory boards; advocacy organizations; and on DOD/DOE panels.

Carolyn Compton, M.D., Ph.D.

comptcar@mail.nih.gov

Carolyn Compton, M.D., Ph.D., is the Director of the Office of Biorepositories and Biospecimen Research (OBBR) at the NCI. In these capacities, she has leadership responsibility for strategic initiatives that include the Cancer Human Biobank (caHUB) project, the Innovative Molecular Analysis Technologies for Cancer program, and the NCI Community Cancer Centers project. She is an adjunct Professor of Pathology at the Johns Hopkins School of Medicine. She received her M.D. and Ph.D. in degrees from Harvard Medical School and the Harvard Graduate School of Arts and Sciences. She is trained and boarded in both Anatomic Pathology and Clinical Pathology at Harvard's Brigham and Women's Hospital. She came to the NCI from McGill University where she had been the Strathcona Professor and Chair of Pathology and the Pathologist-in-Chief of McGill University Health Center from 2000-2005. Prior to this, she had been a Professor of Pathology Harvard Medical School and the Massachusetts General Hospital, where she was the Director of Gastrointestinal Pathology for 15 years. Dr. Compton has held many national and international leadership positions in pathology and cancer-related professional organizations. Currently, she is the Chair of the American Joint Committee on Cancer (AJCC) and a member of the Executive Committee of the Commission on Cancer. She has published more than 500 original scientific papers, reports, review articles, books and abstracts.

Paul Courtney

paul.courtney@nih.gov

Paul Courtney has been an active contributor in CTMS, TBPT, ICR and DSIC at various times over the last 5 years. While at Dartmouth he managed the Q5 project in ICR, assisted with initial requirements gathering and testing for caTISSUE, participated in discussions in DSIC and was active in CTMS, especially with Patient Study Calendar. He has also been active in the Population Science SIG from the first meeting at the April 2005 caBIG® meeting. He led this group after Eric Ross' tenure and continued to lead when he left Dartmouth to work at Booz Allen Hamilton March 2008. While at BAH he also led the

effort to assist cancer centers in their caBIG® deployments. He joined SAIC in August 2009 to support the PopSciGrid and GEM projects at DCCPS centered in the Behavioral Research Program. In May 2010 added the role of Biomedical Informatics Coordinator for all of DCCPS to his duties. He continues to participate in the Pop Sci SIG, though at a reduced level of effort. Paul is a co-author with other SIG participants from both the NCI as well as cancer centers of "Bioinformatics: Tools to Accelerate Population Science and Disease Control Research" in *Forman et al. Am J Prev Med 2010;38(6):646–651*

Brian Davis, Ph.D.

bdavis@3rdmill.com

Brian Davis, Ph.D., received his Ph.D. in biochemistry and has published research papers on tissue specific gene expression in transgenic mice, pancreatic tumorigenesis in transgenic mice, chromatin structure, mitochondrial gene expression, metabolic engineering, knowledge engineering and ontology best practices. Since 1998, he has been working as a bioinformatacist in various specialties. He joined 3rd Millennium, Inc. in 2005 as the caBIG® Vocabularies and Common Data Elements Workspace (VCDE WS) Lead. Within the caBIG® program, Dr. Davis has been involved in projects for standards development and adoption, enterprise architecture descriptions and recently has been involved in adoption of Services Aware Interoperability Framework (SAIF) within NCI CBIIT and caBIG®. Prior to joining 3rd Millennium, Inc., Dr. Davis was a director at Proteome, Inc., as director of database development and in the area of scientific information aggregation, publishing, product development and software development. Later, he was director of business development at Incyte Genomics, Inc.

Sherri de Coronado, M.S., M.B.A.

decorons@mail.nih.gov

Sherri de Coronado, M.S., M.B.A., is currently acting director, Semantic Services, CBIIT, and government sponsor for the caBIG® VCDE cross-cutting workspace. Trained in molecular biology, biophysics, and science management, Sherri has been part of the Enterprise Vocabulary Services group since its inception in 1997. She helped initiate NCI Metathesaurus, then the NCI Thesaurus-the stand-alone reference terminology for NCI and its collaborators, and co-manages the LexEVS/ CTS2 deployment project. Current interests include open source terminology services, and easy to use tooling to support collaborative ontology building.

Brenda Duggin, R.N., M.S.N.

dugganb@mail.nih.gov

Brenda Duggan, R.N.,M.S.N., is the National Cancer Institute (NCI) Community Cancer Center Program (NCCCP) Information Technology Program Manager. As the IT Program Manager for the NCCCP program she is responsible for supporting the 30 organizations participating in this program, the work of the program, and to support program management activities. Brenda is a Registered Nurse with over 20 years of experience supporting both pediatric and adult oncology. Her role within NCI Center for Biomedical Informatics and Information Technology (CBIIT) is to support the deployment needs of the community-based care setting within caBIG® and Big Health. She is a graduate of Boise State University with a B.S.N. in 1990.

William Dyer

dyerwt@mail.nih.gov

William Dyer has been in the industry for over 18 years, and is a Principal Consultant for Pyramed Research, a professional services company focused on improving the interaction between humans and technology systems used in clinical and translational research networks by promoting the harmonization of standards. He is responsible for providing consulting services, as well as strategic direction and business development for Pyramed Research, and is acting as the caBIG® Clinical Trials Management Systems Representative at the National Cancer Institute. Bill joined Pyramed Research in April 2008, following eight years at Oracle Corporation as a Principal Sales Consultant and Oracle Health Sciences Applications consultant and trainer. Bill was a co-chair for the Oracle Clinical Users Group Training Focus Group from 2005 - 2008, and has worked with most major pharmaceutical companies and CROs throughout the world. Prior to Oracle, he was a Principal Consultant for ErgoTech Corporation working for eight years on system implementation and training projects for Merck and Astra Zeneca. Prior to ErgoTech, Bill was a Principal Consultant, network engineer and instructor for Expert Systems Corporation.

Anna T. Fernandez, Ph.D.

fernandez_anna@bah.com

Anna T. Fernandez, Ph.D., an Associate with Booz Allen Hamilton, has over 12 years of experience in Clinical and Life Sciences. Her professional work and background includes the fields of cancer biology and medical imaging, working with commercial and clinical institutions. In the past, she led projects at Philips Research North America in medical ultrasound imaging for assessing and translating academic research technology into commercial prototypes for improved breast, abdominal, cardiac and vascular imaging for diagnostic and therapeutic efficacy assessment. This included software prototyping and identifying clinical workflow improvements for these new technology areas for Philips Healthcare. Her research has involved conducting initial studies from modeling/simulation work, ex-vivo tissue validation, animal testing, to pilot clinical studies. She has an M.S. and Ph.D. in Biomedical Engineering from Duke University and a background in cancer biology and medical imaging involving digital signal and image processing. She is currently the caBIG® Tissue Bank and Pathology Tools Workspace Lead for the NCI Center for Biomedical Informatics and Information Technology, and on imaging-related efforts for the Pediatric Heart Network with the New England Research Institute supported by a National Heart Lung and Blood Institute grant.

Ian Fore, D.Phil.

forei@mail.nih.gov

Dr. Ian Fore leads Biorepository and Pathology Informatics at the US National Cancer Institute's Center for Biomedical Informatics and Information Technology. He worked on drug discovery informatics at Wyeth Research and Johnson & Johnson Pharmaceutical R&D including developing global databases for research data. More recently he was a product manager at Celera Genomics responsible for integrating Celera's customer's informatics systems. Prior to leaving the lab for an informatics career Ian gained his D.Phil. in Physiology and worked as a Research Pharmacologist. His current role as Associate Director at NCI-CBIIT includes NCI coordinator for the caBIG® Tissue Banks and Pathology Tools Workspace.

Joe George

joe.george@osumc.edu

Joe George is a Senior Technical Lead in the Center for IT Innovation in Healthcare in the Department of Biomedical Informatics at the Ohio State University. Joe's current assignments give him the ability to develop solutions on grid-based middleware for the Cancer Biomedical Informatics Grid (caBIG®) and support the caGrid user community as an information resource to facilitate collaborative scientific research using caGrid.

Russell E. Glasgow, Ph.D.

Russell E. Glasgow, Ph.D., is the new Deputy Director for Dissemination and Implementation Science in the Division of Cancer Control and Population Sciences at the NCI (Sept. 2010). He is a behavioral scientist who has conducted public health research in community, worksite and healthcare settings, focused on the development and evaluation of practical, generalizable interventions. Dr. Glasgow has published over 400 peer reviewed articles and has received the Society of Behavioral Medicine award for Outstanding Scientist (2000). His interests include translational research (he was formerly Co-Director of the Center for Health Dissemination and Implementation Research (www.research-practice.org) at the Institute for Health Research, Kaiser Permanente Colorado); health communication, especially using multiple, new and interactive technologies; and use of conceptual models for pragmatic research to help design, implement, evaluate, and report on studies to accelerate translation of research to practice and policy (www.re-aim.org)

Shannon Hastings, M.S.

shannon.hastings@inventrio.com

Shannon Hastings, M.S., is a Co-Founder of Inventrio. He is experienced in healthcare data integration platforms, distributed and high performance computing, large-scale semantic and syntactic data integration, software development methodologies, enterprise architecture, scientific data visualization, and project management. His recent accomplishments include the creation of the openMDR; a project utilizing the cgMDR from the UK Cancergrid in order to provide localized institutional support for semantically annotated data models in caGrid services, and the Introduce Grid Service Authoring Toolkit, an open source project aimed at providing non-grid and web services computing experts the ability to easily and quickly create secure grid services. Previously Shannon served as co-Director of the Software Research Institute (SRI) in the Center of IT Innovations in Health Care (CITIH) at the Ohio State University. Prior to that he worked as a Research Scientist at General Electric's Global Research Center in the Visualization and Computer Vision and Advanced Computing Technologies Laboratories. Shannon received his M.S. in Computer Science from Rensselaer Polytechnic Institute and B.S. in Computer Science from the Ohio State University.

Mervi Heiskanen, Ph.D.

heiskame@mail.nih.gov

Mervi Heiskanen, Ph.D., is the Associate Director for Community Alliances for NCI CBIIT. Dr. Heiskanen provides guidance and support to the research community on available cancer Biomedical Informatics Grid® (caBIG®) resources. Dr. Heiskanen

received her Ph.D. in human genetics from the University of Helsinki, Finland. Prior to joining the NCI, Dr. Heiskanen was developing high throughput array based technologies as a Senior Scientist at Compugen Inc. and a Visiting Fellow at the NIH Human Genome Research Institute.

Matthew Holt, M.A., M.S.

matthewseparate@gmail.com

Matthew Holt, M.A., M.S., has spent more than 15 years in health care as a researcher, generalist forecaster, and strategist. He's conducted in-depth studies about the health care market, information technology and policy for public release and private clients. He learned from some of the best in forecasting, policy and survey organizations, but these days he's best known as the author of *The Health Care Blog* (http://www.thehealthcareblog.com), and as the co-founder of the *Health 2.0 Conference*. For that he's been mostly self-taught! Matthew started his health care career conducting international health policy research at Stanford's Asia Pacific Research Center. At the Institute for the Future, the Menlo Park, CA think-tank, he led projects in healthcare financing, delivery and information technology. At Harris Interactive, the New York-based survey research firm, Matthew conducted two landmark survey research studies, *Computing in the Physician's Practice* and *The 10,000 Patients Study*. In his last real job back before 2003, Matthew was Vice President, Strategy and Business Development at i-Beacon.com, an eHealth software company. He holds an M.A. in Political Science and an M.S. in Health Services Research from Stanford University.

Jim Humphries

jbh64@georgetown.edu

Jim Humphries has 19 years of experience in Information Technology. His expertise includes software engineering and scientific computing. Since May, 2010, Mr. Humphries is leading the development efforts for cancer Bench-to-Bed application at Georgetown University. Prior to joining the Georgetown University team he worked with Emory University on caMicroscope. Since 1998, he has led QuaTeams, an IT consulting firm. Clients include DoD, NOAA, National Geographic, Peace Corps, and Bacardi. Mr. Humphries holds a degree in computer science from the University of Maryland, College Park.

Amy K. Jacobs, M.S.N., RN-BC

amy.k.jacobs@lmco.com

Amy K. Jacobs, M.S.N., RN-BC, is a Deputy Project Manager at Lockheed Martin Corporation. She currently provides medical expertise for the National Library of Medicine (NLM) Unified Medical Language System (UMLS) and the National Cancer Institute (NCI) Enterprise Vocabulary Services (EVS). Ms. Jacobs has 24 years of healthcare-related experience including direct patient care in clinical settings as a registered nurse and family nurse practitioner as well as healthcare informatics experience in various informatics positions.

Conrade Carl Jaffe, M.D.

carl.jaffe@bmc.org

Conrade Carl Jaffe, M.D., is currently a Professor of Radiology, Boston University and Professor Emeritus of Medicine (Cardiology), Yale University School of Medicine. He holds a B.S. from the Massachusetts Institute of Technology and an M.D. from Columbia College of Physicians and Surgeons and is board

certified in diagnostic radiology and nuclear medicine. Recently he served as Branch Chief for clinical imaging in the Cancer Imaging Program of the extramural division of the National Cancer Institute at NIH from 2003 to 2008. His career focus has been: research in computer applications in diagnostic imaging; education (network-based and classroom); and clinical trial imaging. He has been a visiting scientist at the Office of High Performance Computing and Communications at the National Library of Medicine and founded the Center for Advanced Instructional Media at the Yale School of Medicine and was a recipient of the Pirelli International (sic) Award in 2005 for communication of science and technology carried out entirely on the Internet. He has served as a Board Member of the Whitaker Foundation, Chairman of the Board of Scientific Counselors National Library of Medicine and 2nd Vice President of the Radiological Society of North America. His grants have included an NIH Research Career Development Award, James Picker Foundation Scholar, and an R01 grant from the National Institutes of Health. He has published or edited 8 books and more than 180 monographs and scientific journal articles.

Thomas (Tom) Jones, M.D.

tom.jones@tolvenhealth.com

Thomas (Tom) Jones, M.D., is a founding partner and Chief Medical Officer of Tolven Inc., a start-up company dedicated to furthering the development and use of open source software for healthcare. Following his graduation from medical school at Stanford University, Dr. Jones joined the University of Chicago residency program and was a member of the Department of Medicine faculty until 1995. As part of the development of the infrastructure for primary care education and clinical activity, Dr. Jones and his colleagues at the University developed the Centennial Patient Care Workstation, a model for allowing clinicians to enjoy the benefits of new information technology. His interest in clinical informatics grew out of both his clinical practice experience and his teaching experience. In 1995, Dr. Jones joined Oacis Healthcare Systems where his role allowed him to focus more deeply on the clinical functionality of applied informatics. During his 5 years at Oacis, he had the opportunity to work closely with some of the founding members of the HL7 organization. In 2000, Dr. Jones joined Oracle where, as Chief Medical Officer, he provided the clinical leadership for Oracle's Healthcare Strategy group, including the development of Oracle's Healthcare Transaction Base. In 2002, Dr Jones joined a European Commission Task Force whose work culminated in a directional paper on Biomedical Informatics in 2003. In 2004, Dr. Jones began his participation in the Interoperability Consortium (IC), where he chaired the Technical Committee of the IC and was responsible for the technical and architectural sections (including the discussions of standards) of the IC's response to the ONCHIT RFI. Dr Jones left Oracle in 2006 to found Tolven. He currently chairs the Principles Committee of the Health Record Banking Alliance and the HIMSS HIE Open Source Task Force. He is a member of the ONC's Technical Expert Panel for the Study and Report on Open Source Health Information Technology Systems.

Michael Keller, Ph.D.

keller_michael@bah.com

Dr. Keller, a Senior Associate with Booz Allen Hamilton, has worked in the caBIG® program for 7 years and is currently the caBIG® Architecture Workspace Lead. Dr. Keller, in conjunction with the VCDE WS Lead, coordinates the activities of the Architecture and VCDE Workspaces. These Workspaces are responsible for developing the standards based grid infrastructure that provides the framework for semantic and

syntactic interoperability across disparate data and analytical systems that are part of caBIG®. Prior to joining Booz Allen, Dr. Keller performed research in the field of molecular virology for eight years. Dr. Keller received his Ph.D., in Microbiology/Immunology from Wake Forest University.

Karen A. Ketchum, Ph.D.

Karen.Ketchum@esacinc.com

Karen A. Ketchum, Ph.D., is a Director for Enterprise Solutions And Consulting, E-SAC, a professional services company specializing in data management and bioinformatic solutions. Dr. Ketchum has worked in biotechnology for the past 15 years and has experience in genomic sequencing, disease association studies, and companion diagnostics. She is currently working as a Subject Matter Expert on the caIntegrator project team at the National Cancer Institute (NCI). Prior to joining E-SAC, Dr. Ketchum was Associate Director for Alliance Management at Celera where she supported business development initiatives in pharmacogenomics. Dr. Ketchum received her Ph.D. from McGill University in Biology and continued her laboratory research training as a postdoctoral fellow in the Department of Genetics at Yale University School of Medicine.

Jeff Kiefer, Ph.D.

jkiefer@tgen.org

Jeff Kiefer, Ph.D., is an Investigator at the Translational Genomics Research Institute. His research activities involve utilizing bioinformatic tools and resources to enable knowledge recovery from disparate genomic data. His work at the Translational Genomics Research Institute is to assist individual investigators in translational informatics and knowledge mining specifically in the area of oncology and precision medicine.

Juli Klemm, Ph.D.

klemmj@mail.nih.gov

Juli Klemm, Ph.D., is the Associate Director of Integrative Cancer Research Products and Programs at the NCI Center for Bioinformatics and Information Technology (NCI CBIIT). Dr. Klemm is the Product Manager for a number of ICR tools and is the Facilitator of the caBIG® ICR Workspace. Dr. Klemm's interests are in creating end user-focused systems for data management and analysis that support life science research. Prior to joining NCI CBIIT, Dr. Klemm was the Head of Scientific Consulting at 3rd Millennium, Inc., where she was the ICR Workspace lead throughout the pilot of the caBIG® program. She earlier served as Associate Director of Scientific Project Management at Incyte Corporation. Dr. Klemm received her Ph.D. in Biology from the Massachusetts Institute of Technology for crystallographic and biochemical studies of protein-DNA interactions and did her post-doctoral research at Stanford University in the area of T-cell signaling.

Dan Kokotov, M.S.

Dan Kokotov, M.S., is the Director of Research & Development for 5AM Solutions, Inc. He brings over eight years of experience in software development and a proven track record of success throughout his career as a technical leader and innovator. In this current role at 5AM, Dan leads adoption of new technologies and ensures 5AM stays at the cutting edge of science and technology. Prior to joining 5AM, Dan served as the Director of Software Engineering at Vecna Technologies, Inc., overseeing the company's development processes and cross-product design. A staunch supporter

of open source technology, Dan is the primary author of the XSnapshot open source project. Dan holds a Masters of Engineering in Computer Science and a Bachelor of Science in Computer Science and Mathematics from MIT, where his graduate work involved research in lightweight formal methods and distributed computing.

George Komatsoulis, Ph.D.

komatsog@mail.nih.gov

George Komatsoulis, Ph.D., is Deputy Director of the NCI Center for Biomedical Informatics and Information Technology (CBIIT) where he has broad responsibility for activities within CBIIT and the caBIG® program. He is also the Chief Information Officer (acting) of the National Cancer Institute. He holds a Ph.D. in molecular biology and biochemistry from the California Institute of Technology, and did postdoctoral work in the biochemistry department at the Johns Hopkins University School of Medicine and in the mathematics department at the University of Southern California. Prior to joining NCI, Dr. Komatsoulis was a Senior Bioinformatics Scientist at Human Genome Sciences, in Rockville, MD.

Stephen Langella

stephen.langella@inventrio.com

Stephen Langella is one of the Co-Founders of Inventrio. Stephen brings significant leadership and expertise in enterprise architecture, software development, and security. His accomplishments include architecting and leading the development of a national security infrastructure (GAARDS), which serves as the security infrastructure for several large scale national networks. He also served as a technical architect, security architect and senior developer for the caGrid project, a national platform for health care integration. caGrid/GAARDS provides the software foundation for the Cancer Biomedical Informatics Grid (caBIG®), a national initiative funded by the National Cancer Institute (NCI). Mr. Langella has served in a leadership position and has made significant contributions to caBIG® since its inception. caGrid/GAARDS has been widely adopted by other large national efforts, including the Clinical Translational Research Awards (CTSA), Cardiovascular Research Grid (CVRG), the Biomedical Informatics Research Network (BIRN), and others. Previously, Mr. Langella co-founded and co-directed the Software Research Institute (SRI) in the Center of IT Innovations in Health Care (CITIH) at the Ohio State University, wherein he helped coordinate the center's software development and research projects towards the aim of producing high-quality reusable infrastructure for both national and internal use. Under Mr. Langella's leadership SRI/CITIH grew from a small four man research team to a nationally recognized center for informatics research. Mr. Langella received his B.S. in Computer Science from the University of Buffalo and his M.S. in Computer Science from Rensselaer Polytechnic Institute.

Caterina Lasome, Ph.D., M.B.A., R.N.

lasomece@mail.nih.gov

Caterina Lasome, Ph.D., M.B.A., R.N., serves as the Chief Operating Officer for the National Cancer Institute's (NCI) Center for Biomedical Informatics and Information Technology (CBIIT). She has broad responsibility for the programmatic, financial, contractual oversight, and software development life cycle for NCI's Cancer Biomedical Informatics Grid (caBIG®) program, a virtual network of interconnected data, individuals, and organizations that redefines how research is conducted, care is provided, and patients/participants interact with the biomedical research enterprise. Dr. Lasome

further serves as the Executive Government Sponsor for NCI's cancer Electronic Health Record (caEHR) effort, an ARRA-funded initiative to develop specifications for an "oncology-extended Electronic Health Record" to fill a gap in existing Electronic Health Record (EHR) functionality for practicing ambulatory oncologists. Additionally, Dr. Lasome serves in the capacity of Advisor for the Johns Hopkins University HITECH-funded Curriculum Development Center Grant and the National Quality Forum's Health Information Technology Advisory Committee.

Zhong Li, Ph.D.

zli@c2b2.columbia.edu

Zhong Li, Ph.D., Director of Operations, is currently the Operation Manager of the caBIG® Molecular Analysis Tools Knowledge Center (MAT-KC). The MAT-KC supports four biomedical data management and analysis tools: caArray, caIntegrator/caIntegrator2, geWorkbench, and GenePattern. Dr. Li received his Ph.D. in Biochemistry, Molecular Biology, and Genetics from the Johns Hopkins University School of Medicine. He received his M.S.-level training in Computer Science from the Johns Hopkins University School of Engineering. Before coming to Columbia to manage the MAT-KC, he worked in both large pharmaceutical companies such as GlaxoSmithKline and small biotech start-up companies for more than 10 years with various roles and responsibilities. His areas of expertise include Bioinformatics data management/analysis, enterprise software development, algorithm development, and project/operation management.

Jack London, Ph.D.

Jack.London@KimmelCancerCenter.org

Jack W. London, PhD, is a Research Professor of Cancer Biology in the Jefferson Medical College, Thomas Jefferson University, and is Director of Jefferson's Kimmel Cancer Center's Informatics Shared Resource. He has been engaged in research and system development in the domains of bioinformatics, medical informatics, and health care information technology for over 35 years. This work has encompassed many areas: computer simulation of biochemical systems (including computer-aided design of clinical laboratory enzyme assays), radiology information system (RIS) and picture archiving and communications system (PACS) development, hospital information systems development, telemedicine, clinical trials management systems (including adverse event reporting), and tissue banking applications. The focus of his work in recent years has been on biospecimen and clinical trials informatics, with emphasis on collaborative development of interoperable systems. Such efforts included work with the Pennsylvania Cancer Alliance Bioinformatics Consortium (PCABC) beginning in 2002 and projects with the National Cancer Institute's "cancer Biomedical Informatics Grid" (caBIG®) initiative since its inception in 2004. The caBIG® activities included an ICR caArray project, TBPT projects with caTIES and caTissue, CTMS projects with Patient Study Calendar and the Protocol Lifecycle Tracking tool, and participation with the DSIC workspace. Dr. London has also served on a number of National Institutes of Health grant review panels.

Susan Love, M.D., M.B.A.

Susan M Love, M.D., M.B.A. has dedicated her professional life to the eradication of breast cancer. As President of the Dr. Susan Love Research Foundation, she oversees an active research program centered on breast cancer cause and prevention. She is also a Clinical Professor of Surgery at UCLA's David Geffen School of Medicine. Her reputation as an activist comes from her role as one of the "founding

mothers" of the breast cancer advocacy movement in the early 1990's by being one of the founders of the National Breast Cancer Coalition (NBCC). She continues this work by serving on the boards of the National Breast Cancer Coalition, and Young Survival Coalition. She served on the National Cancer Advisory Board from 1998-2004 as an appointment of President Clinton. Dr. Susan Love is best known as a trusted guide to women worldwide through her books and the Foundation website. The completely revised fourth edition of *Dr. Susan Love's Breast Book* termed "the bible for women with breast cancer" by The *New York Times*; was released October 2005, and the 5th edition will be coming out in 2010. It has been translated into German, Dutch, Chinese, Japanese and Hebrew. *Dr. Susan Love's Menopause and Hormone Book*, first published in 1998 and revised in 2003, was one of the first to sound the alarm against the long term use of postmenopausal hormones and *Live a Little* (Crown 2009) encourages women to be healthy without driving yourself crazy. A true visionary, Susan Love's most recent project, the Love/Avon Army of Women, is a creative Internet solution to partner women and scientists in order to accelerate basic translational research. Thanks to a generous grant from the Avon Foundation for Women, the Dr. Susan Love Research Foundation launched the Army of Women campaign in October 2008. The campaign is recruiting one million women who are willing to consider participating in research to find the cause and prevention of breast cancer. This novel initiative will move research from animals to women, democratizing the scientific process. Dr. Love received her medical degree from SUNY Downstate Medical Center in New York, did her surgical training at Boston's Beth Israel Hospital. She founded the Faulkner Breast Center in Boston and the Revlon UCLA Breast Center in Los Angeles. She has a business degree from the Executive MBA program at UCLA's Anderson School. In 1996 she retired from the active practice of surgery, to dedicate her time to the urgent pursuit of finding the cause and prevention of breast cancer. The Dr. Susan Love Research Foundation, a 501 (c) 3 non-profit breast cancer foundation has over $4 million dollars in peer reviewed grants and is pioneering novel techniques to identify young women at risk for breast cancer as well as local therapy directly into the milk ducts to prevent breast cancer.

Cecil O. Lynch, M.D., M.S.

clynch@ontoreason.com

Cecil O. Lynch, M.D., M.S., is an assistant Professor at University of California Davis school of medicine in the division of pathology informatics, and also the CEO and chief knowledge engineer for an artificial intelligence medical software company called OntoReason He is also a national science advisor to both the US Centers for Disease Control national Center for public health informatics and the European Centers for Disease Control in knowledge management. He is on the board of directors for the BRIDG Model at National Cancer Institute and on the architecture review board for HL7. He is presently the Chief Semantic Architect at NCI CBIIT.

Ravi K. Madduri

madduri@mcs.anl.gov

Ravi K. Madduri is a Project Manager in the Mathematics and Computer Science Division at Argonne National Laboratory and an honorary research fellow at the Computation Institute at the University of Chicago. Ravi is one of three key contributors to the National Institutes of Health $100M Cancer Biomedical Informatics Grid (caBIG®), which links 60 NIH-funded cancer centers and clinical sites engaged in cancer research. For his efforts in project management, tool development, and collaboration, Ravi received several Outstanding Achievement Awards from NIH in recognition of his work on caBIG® project

management, tool development, and collaboration. Ravi is a lead architect on the scientific workflow design and implementation project under the caGrid toolkit.

Subha Madhavan, Ph.D.

sm696@georgetown.edu

Subha Madhavan, Ph.D., is an information scientist who has worked in the field of biomedical informatics and clinical data management and analysis for the past 11 years. Prior to joining Georgetown, Dr. Madhavan served as the Associate Director of Product and Program Management in the Life sciences informatics area at the NCI's Center for Biomedical Informatics and Information technology (CBIIT). Her work at the NCI involved bridging the gap between bench and bedside by enabling researchers and physician scientists to use cutting edge biomedical informatics solutions to identify better therapies for cancer. Specifically, she led a group of scientists, physicians and software engineers in building REMBRANDT (REpository for Molecular BRAin Neoplasia DaTa) – a database that hosts and interconnects clinical data points with various genomics datasets from large brain tumor clinical trials. At Georgetown, she leads various informatics efforts including the caBIG® *In Silico* Research Center of Excellence, Cancer Family Registries and the Georgetown Database of Cancer. She also directs the Informatics cores for NIH funded translational efforts such as the CTSA and the Center for Cancer Systems Biology at Georgetown. More recently, she has joined forces with BigHealth initiatives in integrating electronic patient reported outcomes with clinical information to advance breast cancer care and research.

Joshua Mann

Josh.Mann@stjoe.org

Josh Mann has over 13 years experience in the IT industry, 5 of which have been in the healthcare vertical. He has actively sought out ways to enhance and expand the use of technology within healthcare ranging from development of tools for direct caregiver support to the integration of systems across hospital, private practice and government agency sectors. His experience ranges from developing software for medical devices to the oversight for the implementation of an electronic medical records system. Josh continues to oversee the implementation of information technology systems for the SJO Cancer Center.

Melissa Markey, J.D.

mmarkey@hallrender.com

Melissa Markey, J.D., is licensed to practice as an attorney in Texas and Michigan and currently works for the law firm of Hall, Render, Killian, Heath & Lyman, PLLC. Ms. Markey primarily represents health care providers, exempt organizations, and health information exchange entities. Ms. Markey has a particular interest in technology law, including electronic health records, software licensing, data rights management and e-commerce issues, as well as emergency legal preparedness, HIPAA, patient care issues, human subjects research, and compliance. Ms. Markey has presented and authored publications both within Michigan and nationally on topics including legal and policy issues in health information exchange, privacy and security under HIPAA and other laws, emergency preparedness and response law, the clinical-technology interface, and clinical research issues. Melissa is on the Board of Directors of the American Health Lawyers Association, as well as the Michigan Chapter of Healthcare Information and

Management Systems Society ("HIMSS"). Ms. Markey is also a Michigan licensed and Nationally Registered Emergency Medical Technician-Paramedic with many years of field experience.

Cheryl Marks, Ph.D.

marksc@mail.nih.gov

Cheryl Marks, Ph.D., is Associate Director of National Cancer Institute's Division of Cancer Biology; she designs and implements programs to integrate basic cancer research discoveries into translational, clinical, and population science. To augment support of the Institute's basic science programs, she identifies private sector and foundation partners, and coordinates trans-NCI activities for preclinical model systems with similar programs across the National Institutes of Health and other Federal agencies. Dr. Marks has held her present position since 2000. She earned a bachelor's degree in chemistry from Randolph-Macon Woman's College and a master's degree in chemistry from Southern Connecticut State University. Her doctorate in biochemistry and genetics is from the George Washington University. In 1999, Dr. Marks implemented an innovative, interdisciplinary NCI program, the NCI-Mouse Models of Human Cancers Consortium, an international assembly of at least 300 basic and clinical cancer researchers from more than 70 institutions in the US and abroad. The exceptionally diverse expertise of the participants ensures the success of the Consortium in generating mouse cancer models that simulate the natural history and clinical course of the corresponding human diseases, and in pioneering the innovative systems biology approaches that enable their use for translational research.

Karen J. Maschke, Ph.D.

maschkek@thehastingscenter.org

Karen J. Maschke, Ph.D., is a Research Scholar at The Hastings Center and the Editor of IRB: Ethics & Human Research. Maschke has been active in research and public policy initiatives related to ethics and human subjects research. She was a project member on two NHGRI/ELSI-funded projects: "Ethical Decision-Making for Newborn Genetic Screening," and "Education in Genetic Ethics (EDGE)." She currently is a collaborator on the NIDA-funded project, "Law & Ethics of Drug Addiction Genetics Research (LEDGER)." Maschke was a participant at the NCI Workshop on Biospecimen Access and Ethical, Legal, and Policy Issues, the NCI/NHGRI Data Release Workshop for The Cancer Genome Atlas (TCGA), and the NCI Symposium on Custodianship and Ownership Issues in Biospecimen Research. She served recently as the co-chair of the Ethics Subgroup for the NCI's Cancer Human Biobank (caHUB). Maschke currently is funded by the NCI to assist the Data Sharing and Intellectual Capital work space (DSIC) of the NCI's Cancer Biomedical Informatics Grid (caBIG®) in developing policies and guidelines for researchers and Institutional Review Boards regarding access controls to cancer research data and to assist the DSIC Knowledge Center in developing support tools pertaining to the ethical, regulatory, and policy issues for data sharing, data integration and information exchange via the caBIG® infrastructure. Maschke is a member of the IRB at Vassar Brothers Medical Center in Poughkeepsie, New York.

Jim McCusker, M.S.

james.mccusker@yale.edu

Jim McCusker, M.S., is a software developer at Yale School of Medicine, where he is the lead developer for the Yale SPORE for Skin Cancer. He has contributed code for importing Nimblegen micrarray

data into caArray. He has served as the Architecture Guide to Mentors for the Tissue Banking and Pathology Tools Workspace, and is investigating generic models of provenance for caBIG®. He is also a Ph.D. student in Computer Science at Rensselaer Polytechnic Institute.

Charles Mead, M.D., M.Sc.

meadch@mail.nih.gov

Charles Mead, M.D., M.Sc., has over 30 years of experience in digital signal processing and algorithm development, complex software systems and architectures, and healthcare and life sciences informatics. Dr. Mead has experience in clinical trials methodologies and data management systems, application of the Unified Process, and fundamental healthcare and life sciences informatics issues including terminology management, application of the Health Level Seven (HL7) Reference Information Model (RIM), use of Clinical Data Interchange Standards Consortium (CDISC) standards such as SDTM and ODM, the JANUS data model, and Oracle's HTB development framework. Dr. Mead currently is Chair of the HL7 Architecture Board, Chair of the Open Health Tools Architecture Project Team, and a member of the CDISC Board of Directors.

Martin Morgan, Ph.D.

mtmorgan@fhcrc.org

Martin Morgan, Ph.D., leads the Bioconductor project for statistical analysis and comprehension of high throughput genomic data. He is the Director of the Computational Biology Shared Resource at the Fred Hutchinson Cancer Research Center, and has participated in caBIG® activities for several years.

Richard P. Moser, Ph.D.

moserr@mail.nih.gov

Richard P. Moser, Ph.D., is a Research Psychologist within the Office of the Associate Director, Behavioral Research Program, Division of Cancer Control and Population Sciences (DCCPS) at the NCI. He currently functions as the data coordinator for the Health Information National Trends Survey (HINTS), provides analytic support for in-house research projects, serves as the liaison between the cancer prevention fellowship and DCCPS, and performs his own research. He is currently leading a project to create a web-based database, using grid technology, to promote the use of standardized social science measures and data sharing. He also serves on the DCCPS evaluation team that has created new methods and metrics for assessing the value of large scientific initiatives that the division supports. He was a special editor and contributed to several chapters for a supplement to the American Journal of Preventive Medicine regarding the Science of Team Science. His research interests include statistical methodology, health cognitions, and end-of-life issues. Before joining the NCI, he worked at the Palo Alto VA hospital performing alcoholism research, taught statistics at several Bay Area psychology graduate schools and consulted for the statistical software company SPSS. He obtained his doctorate in Clinical Psychology from the Pacific Graduate School of Psychology in 1996.

Uma Mudunuri

mudunuriu@mail.nih.gov

Uma Mudunuri has been with the Advanced Biomedical Computing Center in the Bioinformatics Support Group (BSG) for more than 5 years. She currently heads the Core Infrastructure Group in BSG. She works with complex biological data and strives to offer simple, flexible solutions for challenging data analysis problems.

Sorena Nadaf, M.S., M.M.I.

adamicht@cc.ucsf.edu

Sorena Nadaf, M.S., M.M.I., was recently recruited to serve as Director of the Translational Informatics Program (TIP) at the University of California San Francisco Helen Diller Family (HDFCCC) Comprehensive Cancer Center and Cancer Research Institute. As Principle Investigator, Sorena provides the leadership and vision for the TIP, in addition to oversight of operations and production management of existing biomedical and clinical systems, ensuring their availability, integrity and data services. Also serving as Chief Informatics Officer (CIO) for the UCSF HDFCCC, has allowed Sorena to expand on the creation of a Translational Bioinformatics Shared Resource for clinical and research data. The implementation of the TIP at UCSF has enabled the Helen Diller Family Comprehensive Cancer Center to optimize exciting biomedical discoveries and move the Center more effectively and efficiently toward the ultimate goal of translating discovery into improved clinical outcomes and care delivery. Sorena Nadaf serves as UCSF's liaison to the National Cancer Institute caBIG® Program, but also internally representing the Cancer Center to the institutions Clinical and Translational Science Institute. UCSF was one of the original recipients of the CTSA award. Sorena is intricately involved in the caBIG® Clinical Trial Management Systems Workspace where he is the PI of the "Transcend" software development project, the caBIG® development project that will result in a syntactically and semantically harmonized interoperable patient clinical research tool. He is a proponent of Agile software development methodologies and the Agile Unified Process Framework and the application of these to biomedical research problems.

Rakesh Nagarajan, M.D., Ph.D.

rakesh@wustl.edu

Rakesh Nagarajan, M.D., Ph.D., is an Associate Professor in the Department of Pathology & Immunology at the Washington University (WU) School of Medicine. The long term goals of his lab are to develop and leverage informatics tools in order to uncover perturbations in important pathways leading to disease, resulting in the development of diagnostic tools or the formulation of personalized therapies. Dr. Nagarajan is also co-director of the Alvin J. Siteman Cancer Center (SCC) Bioinformatics Core, director of the Neuroscience Blueprint Biomedical Informatics Core, co-director of the Center for Kidney Disease Research (CKDR) Translational Research Core, and director of Washington University's CTSA Biomedical Informatics Program (BIP). Under the direction of Dr. Nagarajan, separate informatics groups as well as others have been brought together through the establishment of the WU Center for Biomedical Informatics (CBMI). Dr. Nagarajan, collaborators, and his group are implementing a common informatics infrastructure to support the diverse needs of physician-scientists and bench researchers. These are summarized as follows: developing and deploying caTissue Suite institution-wide; a caBIG®-compatible clinical data forms builder, management, and query system, termed ClinPortal, is being developed to facilitate clinical and translational research; developing caBench-to-Bedside (caB2B), which may be used to query one or more caGrid data services and may be used to analyze this data using caGrid analytical services. Dr. Nagarajan and the SCC Bioinformatics Core have also created significant infrastructure to manage, analyze, and annotate microarray expression and SNP and mutation profiling data sets.

Rachel Nosowsky, J.D.

rachel.nosowsky@ucop.edu

Rachel Nosowsky, J.D., serves as Principal Counsel to the University of California system. She specializes in the area of research and clinical trials, with a focus on human research protections, FDA regulation, industry-academic relations and related fraud and abuse and conflict of interest matters, reimbursement, health privacy, biobanking, data repositories, compliance, and misconduct. Ms. Nosowsky provides UC's medical centers with regulatory analysis and advice, supports the University's response to government inspections and oversight initiatives, and reviews and assists in the negotiation of contracts and other legal documents. Before joining the University, Ms. Nosowsky worked in private practice and as Assistant General Counsel at the University of Michigan; and from 2008-2010, she served as Co-Director of the caBIG® Data Sharing and Intellectual Capital Knowledge Center. She currently serves as Vice Chair of the American Health Lawyers Association's Teaching Hospitals and Academic Medical Centers Practice Group and as a member of the University Health System Consortium Legal/Compliance Council's steering committee. Ms. Nosowsky publishes and speaks nationally in the areas of human research and health privacy. She is a graduate of Carleton College (B.A., cum laude, 1988) and the University of California, Berkeley (Boalt Hall) (J.D., 1994).

Scott Oster

scott.oster@inventrio.com

Scott Oster, Co-Founder of Inventrio, is a software architect and engineer with extensive experience leading the design and development of large-scale software infrastructure projects. He has over 10 years experience working on successful open source software projects. Most recently, he has been the Chief Architect for the caGrid project, which is a multi-institutional development effort providing the underlying open source grid software infrastructure used by the Cancer Biomedical Informatics Grid (caBIG®). Since caBIG®'s initial inception in late 2003, Mr. Oster has played a leadership role in the design and implementation of its architecture.

Wendy E. Patterson, Esq.

pattersw@mail.nih.gov

Wendy Ehrenkranz Patterson has served since 2000 as a Senior Advisor in the Technology Transfer Center of the National Cancer Institute (NCI), a participating institute of the U.S. National Institutes of Health (NIH). She works primarily with the NCI's Center for Bioinformatics and Information Technology (CBIIT), where she supports NCI CBIIT's acquisition, development, distribution and deployment of software, infrastructure and research resources on intellectual property, technology transfer, and contractual matters; developing and/or negotiating agreements to share research materials and data through centralized consortia and repositories; supporting compliance with regulatory, policy and mission-related requirements for operation of information systems managed by NCI CBIIT; supporting trademark registration and licensing activities associated with programs managed by NCI CBIIT; and overseeing the caBIG® Data Sharing and Intellectual Capital (DSIC) Workspace and DSIC Knowledge Center, the components of the NCI's (caBIG®) program that seek to facilitate appropriate data sharing between and among organizations by addressing legal, regulatory, policy, ethical, proprietary, contractual, and socio-cultural barriers. In addition, she supports NCI extramural scientific program directors and their staffs with respect to data sharing and intellectual property matters arising from proposed grants, contracts and cooperative

agreements for collaborative research networks and consortia.

From 1988 to 1998, Ms. Patterson served as an Attorney-Advisor in the Office of General Counsel of the U.S. Commodity Futures Trading Commission where she was involved in appellate litigation, preparation of final adjudicatory decisions of the Commission and regulatory reviews. Earlier, she was an attorney in private practice at a law firm in Washington, D.C. where she focused on corporate and securities law.

Ms. Patterson received her A.B. in Psychology and Social Relations from Harvard University in 1976 and her J.D. from the University of Miami School of Law in 1983. She is admitted to the Bar in the District of Columbia.

Philip R.O. Payne, Ph.D.

philip.payne@osumc.edu

Philip R.O. Payne, Ph.D., is the Chair of the Department of Biomedical Informatics at The Ohio State University, and Executive Director of The Ohio State University Center for IT Innovation in Healthcare (CITIH). He also serves as the Director of the Biomedical Informatics Program for OSU's CTSA-funded Center for Clinical and Translational Science and Director of Data Management Services with the Biomedical Informatics Shared Resource of the OSU Comprehensive Cancer Center. Dr. Payne received his Ph.D. with distinction in Biomedical Informatics from Columbia University, where his research focused on the design and evaluation of advanced information management platforms for clinical and translational research. Dr. Payne's research portfolio is actively supported by a combination of NCRR, NLM and NCI awards and contracts. Current projects within Dr. Payne's laboratory include: 1) the development of informatics platforms capable of enabling the collection, exchange and integration of distributed, multi-dimensional, and heterogeneous biomedical data and knowledge sources; 2) the use of knowledge engineering methodologies to support high-throughput discovery of novel phenotype-biomarker linkages and complexes in large-scale data sets; and 3) the application of workflow analysis and human-computer interaction design principles in order to enable the optimal use of information systems in biomedical research and clinical care settings.

Justin Permar

Justin.Permar@osumc.edu

Justin Permar is a Director in the Center for IT Innovations in Healthcare. He leads innovative software development activities via technical direction, outreach, staffing, financial analysis and strategic oversight and planning. Mr. Permar currently serves at the Operations Manager for the National Cancer Institute's Cancer Biomedical Informatics Grid (caBIG®) caGrid Knowledge Center (https://cabig-kc.nci.nih.gov/CaGrid/KC). Previously, Mr. Permar was the lead Grid Software Engineer for the Cardiovascular Research Grid (CVRG), an effort funded by the National Heart, Lung, and Blood Institute (http://www.cvrgrid.org). This project aims to facilitate collaborative cardiovascular research by developing and deploying a research Grid that leverages the caGrid middleware developed by, and used in, the NCI's Cancer Biomedical Informatics Grid (caBIG®) program. Mr. Permar joined The Ohio State University after nearly five years of research and development at the Massachusetts Institute of Technology Lincoln Laboratory (MIT LL) in Lexington, MA. His research at MIT LL focused on creating high-performance near-real time middleware that served as the distributed platform for a live radar control center. Mr. Permar received a Bachelors of Science in Computer Engineering with Honors from Brown University in 2002. His honors thesis

research work was part of a project to engineer a brain-machine interface to conduct neuroscience research. His efforts focused on digital signal processing systems designed to capture and analyze neural signals in the motor cortex.

Sheila A. Prindiville, M.D.

Sheila A. Prindiville, M.D., is the Director of the Coordinating Center for Clinical Trials at the National Cancer Institute (NCI). She was appointed to this position in 2006 to oversee the restructuring of the national cancer clinical trials enterprise to ensure that it is not only more efficient and coordinated, but founded on the best science. Dr. Prindiville received her M.D. from Northwestern University and her M.P.H. from Johns Hopkins University. After finishing her residency in Internal Medicine at Evanston Hospital, she completed fellowships in both Medical Oncology and Cancer Prevention at the National Cancer Institute in 1995. She was on the faculty of the Division of Medical Oncology at the University of Colorado Health Sciences Center from 1996 to 2001, where she conducted research in cancer epidemiology and prevention. She returned to the NCI in 2002 and directed the Clinical Cancer Genetics Program in the Genetics Branch of the Center for Cancer Research prior to her current appointment. She is an adjunct staff clinician in the Medical Oncology Branch where she conducts clinical trials developing targeted therapies for individuals with inherited predisposition to cancer.

Ken Quinn, R.N.

ken.quinn@roswellpark.org

Ken Quinn, R.N., has served as the caBIG® Deployment Lead at Roswell Park Cancer Institute since 2008 and is a member of the Information Technology Department. He directs all caBIG®- related activities including strategic planning, implementation and integration efforts. Ken has 29 years of healthcare experience. The past 14 years are in healthcare information technology. Working as a systems analyst and IT project manager, he has played key roles in many large institutional IT initiatives. Currently, he is directing the implementation of a virtual data warehouse at Roswell Park using caGrid technology.

Dianne M. Reeves, R.N., M.S.N.

reevesd@mail.nih.gov

Dianne M. Reeves, R.N., M.S.N., is the Associate Director for Biomedical Data Standards in the National Cancer Institute Center for Biomedical Informatics and Information Technology (CBIIT). An oncology nurse with extensive experience in clinical trials conduct, education, and research, Ms. Reeves has been working with biomedical informatics and data standards for the past decade. Her combination of clinical experience and biomedical informatics expertise allow her to act in the role of liaison and facilitator between clinical research teams and software developers.

Konrad Rokicki, M.S.

rokickik@mail.nih.gov

Konrad Rokicki, M.S., is a Senior Software Engineer at SAIC, Inc. and has been a developer on the caBIO project since 2006. He has also contributed to other caBIG® projects such as EVS and caDSR. He got started with iPhone apps in 2009 with the caBIO iPhone App, and has been working on “mobilizing” other caBIG® data ever since.

Kristen Rosati, J.D.

krosati@csblaw.com

Kristen Rosati, J.D., is a partner in the law firm of Coppersmith Schermer & Brockelman, PLC. Her practice concentrates in clinical research, electronic health records, health information privacy and security, and consent issues. Much of her work is at the intersection of these areas, including data sharing in collaborative research, the creation of data warehouses and tissue banks for research, and "secondary" use of health information. Ms. Rosati is assisting in the development of the Arizona Translational Resource Network (AzTransNet), an effort by the Arizona Biomedical Research Commission to build clinical and translational research capacity in Arizona. AzTransNet is working to serve the needs of research sites across Arizona's biomedical community, including developing the Arizona Biospecimen Consortium (a "virtual" tissue bank). Ms. Rosati also is assisting with the creation of the FDA Sentinel Initiative, an effort to create a national electronic distributed network to monitor medical product safety. She is serving on the Privacy Panel for the "Mini-Sentinel Coordinating Center" at Harvard Pilgrim, is on the Brookings Active Surveillance Implementation Council (BASIC), and served on the planning committee for "Legal Issues in Active Medical Product Surveillance" conference sponsored by the Engelberg Center for Health Care Reform at the Brookings Institution. Ms. Rosati is a member of the American Health Lawyers Association (AHLA) Board of Directors, serves as Secretary on its Executive Committee, and Chairs its Professional Resources Committee. Ms. Rosati received her B.A. and J.D. with honors from the University of Michigan.

Eric A. Ross, Ph.D.

Eric A. Ross, Ph.D., is Assistant Vice President of Biometrics and Information Sciences at Fox Chase Cancer Center (FCCC). He is responsible for directing FCCC's biostatistics, bioinformatics and research informatics activities. Dr. Ross has graduate degrees in both Biostatistics and Statistics. His research focuses on the innovative use of quantitative methods and information technology in cancer research. He is a funded participant as a Subject Matter Expert for the caBIG® Population Sciences Special Interest Group. Before joining FCCC in 1989, Dr. Ross held positions with the University of South Florida and industry.

John Sandefur, M.B.A.

john.sandefur@ccc.uab.edu

John Sandefur, M.B.A., is an Information Systems Manager and the caBIG® Deployment Lead at the University of Alabama at Birmingham Comprehensive Cancer Center. John earned his B.A. degree in Accountancy from the University of Mississippi in 1981, an M.B.A. degree from the University of Nebraska in 1987, and is currently completing an M.S. degree in Health Informatics at the University of Alabama at Birmingham. He has worked in mergers and acquisitions, information technology, and business consulting, as well as owning an information technology business. He has published five journal articles on entrepreneurship.

Thomas P. Sellers, M.P.A.

President and CEO of the National Coalition for Cancer Survivorship (NCCS), Tom Sellers is an 11-year cancer survivor whose life mission is to ensure the best possible quality of life for all cancer survivors. Since his mother's death from lung cancer, he has spent the last 30 years relentlessly pursuing

that mission in both his private and professional life. Mr. Sellers arrived at NCCS in October 2009 and leads this national non-profit organization advocating for quality cancer care that is patient-centered, comprehensive, and coordinated. Before joining NCCS, Mr. Sellers led the fundraising, community relations, design, and construction for a $30 million American Cancer Society (ACS) project to build a 40 suite Hope Lodge in Boston providing free lodging and services to cancer patients in treatment. Opened in 2008, the facility provides free lodging to more than 1,000 cancer patients annually. From 1995 to 2009 Mr. Sellers also served as CFO for the New England Divisions of ACS. He has over 35 years of experience in public policy and non-profit management, including 13 years in senior positions in Massachusetts state government. Mr. Sellers currently is a member of the Commission on Cancer, on the external advisory board for the Moffit Cancer Center Total Cancer Care Program, on the board of the Schwartz Center at Massachusetts General Hospital, and on the advisory board of the LAF Adult Survivorship Center at the Dana-Farber Cancer Institute. Mr. Sellers is a graduate of the Harvard Kennedy School with a master's degree in public administration and attended Amherst College.

Avinash Shanbhag

shanbhak@mail.nih.gov

Avinash Shanbhag is presently the Director of Core Infrastructure Engineering, at NCI Center for Biomedical Informatics and Information Technology (CBIIT). He has led the development and the operations of the caGrid Infrastructure, and the tools supporting the semantics infrastructure at CBIIT. He is the Architecture workspace lead for the caBIG® program and manages the caGrid Knowledge Center. Avinash is the federal sponsor of the caGrid 2.0 Roadmap activity that is coordinating the development of Semantically aware Service Oriented Architecture (sSOA) at CBIIT. Prior to joining NCI, Avinash worked as a senior Software Development Manager at Manugistics, Inc, a publicly listed supply chain software company. Avinash Shanbhag has a Masters degree in Operations Research and Chemical Engineering. He is also a graduate of NCI's Senior Executive Enrichment and Development (SEED) program.

Ganesh Shankar, M.S.

gshankar@iupui.edu

Ganesh Shankar, M.S., is currently the caBIG® Deployment Lead for the Indiana University Simon Cancer Center (IUSCC) and has served in this capacity for the past three years. In addition, he is the Manager for the Advanced IT Core at the Pervasive Technology Institute, and provides access to high performance storage, applications, and computing infrastructure to the IU School of Medicine. He received his first Masters in Biological Sciences from Carnegie Mellon University and his second Masters in Bioinformatics from the State University of New York, Buffalo. Before joining IU, Mr. Shankar worked as a Bioinformatics Developer, and Project Lead at Roswell Park Cancer Institute. As Deployment Lead, he has contributed to improving caBIG® applications at the national level as the caArray Subgroup Lead. He is also a consultant, and co-founder of PowerHouse Proteomic Systems, a biotech startup that produces fluorescently tagged stem cell lines.

Eliot Siegel, M.D., Ph.D.

esiegel@umaryland.edu

Eliot Siegel, M.D., Ph.D., is Professor and Vice Chair at the University of Maryland School of Medicine, Department of Diagnostic Radiology, as well as Chief of Radiology and Nuclear Medicine for the Veterans Affairs Maryland Healthcare

System, both in Baltimore, MD. Dr. Siegel is also responsible for the NCI's National Biomedical Imaging Archive and serves as the Subject Matter Expert Lead of the National Cancer Institute's caBIG® Imaging Workspace. Under his guidance, the VA Maryland Healthcare System became the first filmless healthcare enterprise in the United States. He has written over 200 articles and book chapters about PACS (Picture Archiving and Communication Systems) and digital imaging, and has edited six books on the topic, including Filmless Radiology and Security Issues in the Digital Medical Enterprise. He has made more than 1,000 presentations throughout the world on a broad range of topics involving the use of computers in medicine. He has been named as Researcher of the Year, received multiple awards for innovation, including the Smithsonian award, and was selected as runner up Educator of the Year for Diagnostic Radiology. The readers and editorial board of Medical Imaging have selected Dr. Siegel as one of the top ten radiologists for the past two years. He was symposium chairman for the Society of Photo-optical and Industrial Engineers (SPIE) Medical Imaging Meeting for three years, is currently chair of Publications for the Society of Computer Applications in Radiology (SIIM) and has been honored as a fellow in that organization. He is chairman of the RSNA's Medical Imaging Resource Committee. His areas of interest and responsibility at both the local and national levels include digital imaging and PACS, telemedicine, the electronic medical record, and informatics. Dr. Siegel holds an M.D. from the University of Maryland.

John Speakman

john.speakman@nih.gov

John Speakman serves as Associate Director for Clinical Products and Programs within the Center for Biomedical Informatics and Information Technology (CBIIT), National Cancer Institute, where he leads CBIIT's informatics initiatives in support of clinical research and electronic health records (EHRs) within the cancer Biomedical Informatics Grid (caBIG®) program. He joined NCI in September 2006 from Memorial Sloan-Kettering Cancer Center (MSKCC) in New York City, where he has worked since joining the institution in 1991 from St. Thomas's Hospital in London, UK. John also chairs the Regulated Clinical Research Information Management (RCRIM) workgroup of the Health Level Seven (HL7) standards development organization, charged with developing international data standards for biomedical research. His personal focus area is maximizing the contribution of information technology to the reengineering of the biomedical research process, especially through exploring the seamless incorporation of sources of clinical and biological data.

Rashmi Srinivasa, Ph.D.

rsrinivasa@5amsolutions.com

Rashmi Srinivasa, Ph.D., is a senior software professional with 5AM Solutions. Dr. Srinivasa has 15 years of experience in various roles including managing/leading teams and developing, researching and integrating emerging technologies into comprehensive solutions. She has an in-depth knowledge and keen interest in the health sciences domain and in distributed systems. She has a proven ability to manage multidisciplinary teams in an agile way to ensure successful and timely deployment, while coordinating all facets of a complex project. Rashmi combines industry qualifications (certified ScrumMaster) with academic credentials (Ph.D. in Computer Science) to bring a broad range of skills and expertise to her projects. Her latest experience includes delivery management, project management, requirements analysis and data modeling on NCI projects like caArray and CTRP.

Craig Stancl, B.S.

stancl.craig@mayo.edu

Craig Stancl, B.S., is a Senior Analyst/Programmer in the Division of Biomedical Statistics and Informatics at Mayo Clinic. Mr. Stancl is the Technical Lead for the LexEVS project and co-chair of the Health Level Seven (HL7) Common Terminology Services Release 2 (CTS 2) specification. Mr. Stancl is an active member of the Vocabularies and Common Data Elements workspace (VCDE WS). Mr. Stancl obtained a B.S. in Computer Science. Prior to joining the Mayo Clinic, Mr. Stancl had 11 previous years of experience at International Business Machines (IBM).

William Stephens

william.stephens@osumc.edu

William Stephens is a Senior Biomedical Informatics Consultant in the Center for IT Innovation in Healthcare in the Department of Biomedical Informatics at the Ohio State University where he focuses his development efforts on caGrid and caGrid Knowledge Center. In these roles he enjoys the ability to develop solutions on grid-based middleware for the Cancer Biomedical Informatics Grid (caBIG®) and support the caGrid user community as an information resource to facilitate collaborative scientific research using caGrid.

David Lloyd Steffen, Ph.D.

steffen@bcm.edu

David Lloyd Steffen, Ph.D., is Director of the Biomedical Informatics Group of the Dan L. Duncan Cancer Center at Baylor College of Medicine. He received his Ph.D. from Brandeis University, completed a postdoctoral fellowship with Robert A. Weinberg at the Cancer Center of the Massachusetts Institute of Technology, was a Scientist at the Worchester Foundation for Experimental Biology and an Associate Professor at the University of Massachusetts and is currently a Professor at Baylor College of Medicine (BCM). He has served as President of Biomedical Computing, Inc, the Director of Informatics of the Human Genome Sequencing Center at BCM and Director of the Bioinformatics Research Center at BCM. His primary interests are the use of best practice software engineering tools to support "-omics" and other large scale, collaborative biomedical research and the development of algorithms and pipelines for the integration and mining of large and diverse datasets to extract maximum meaning from these data.

Baris E. Suzek, M.S.

bes23@georgetown.edu

Baris E. Suzek, M.S., is the associate team lead for bioinformatics in Protein Information Resource (PIR) at Georgetown University and a PIR liaison for the Universal Protein Resource (UniProt) consortium. He has more than 10 years of software development experience in scientific computing and bioinformatics. He was the project manager for the caGrid-enablement of PIR project during caBIG® pilot phase and has been an active participant in ICR, VCDE and Architecture workspaces since the inception of caBIG® program. Currently, he is leading the cancer Bench-to-Bed (caB2B) development effort at Georgetown University and has been participating in ICR Information Representation Working Group and HL7 Semantic-aware Interoperability Framework (SAIF) operationalization activities.

Wei Tan, Ph.D.

wtan@mcs.anl.gov

Wei Tan, Ph.D., is a research professional associate at the Computation Institute, University of Chicago and Argonne National Laboratory. He is one of the core developers of caBIG® workflow system, and has received Teamwork Award and Outstanding Poster Award from US National Cancer Institute in recognition of his contribution in it. His research interests include business and scientific workflows, grid and service-oriented computing, Petri nets, social network analysis, etc. He is now involved in multiple healthcare and biomedicine related projects sponsored by NIH, providing service computing and scientific workflow solutions for domain users. He got his Ph.D. from Tsinghua University, China, in 2008. In 2007 he was a graduate Co-op at IBM T. J. Watson Research Center, NY, USA. He has published more than 20 papers in journals, conferences and book chapters. He also serves as program committee member in multiple international conferences and external reviewer for many international journals. Find more from his homepage at http://www.mcs.anl.gov/~wtan/.

Jennifer Tucker, Ph.D.

jtucker@typetalk.com

Jennifer Tucker, Ph.D., has served as the caBIG® Documentation and Training Workspace Lead since 2006, and as a member of the caBIG® Enterprise Support Network Team since it began in 2008. Jennifer is the Consulting Director at OKA (Otto Kroeger Associates). Past and current work includes facilitating a wide range of technical and scientific work groups, conducting leadership development and teambuilding training workshops, leading organizational assessments and strategic planning projects, and managing outreach and communication projects. She holds a BA in Environmental Science from Wesleyan University, an MS in Management from Purdue University, and a PhD in Science and Technology Studies from Virginia Tech. Jennifer is a certified Project Management Professional (PMP), and is the author of the Introduction to Type® Series workbook Introduction to *Type and Project Management*, an overview of how personality type impacts project management and team performance.

Frank (Trey) White III, Ph.D.

frank.white@fkhealth.com

Frank (Trey) White, Ph.D., Senior Vice President at Feinstein Kean Healthcare, has overseen the coordination of communications and marketing activities for the National Cancer Institute's cancer Biomedical Informatics Grid (caBIG®) program for the past 2 years. Frank's background includes almost 15 years of molecular biology and bioinformatics research experience in pharma and contract research labs, along with more than 10 years experience in marketing, communication, alliance management, and business development in the commercial bioinformatics sector. His prior positions included Director of Sales & Marketing for Gene Codes Corporation, a bioinformatics software company located in Ann Arbor, Michigan, program marketing and alliance management positions at both IBM Life Sciences and InforMax, and laboratory research positions at Berlex Biosciences and SRA Technologies.

Raimond L. Winslow, Ph.D.

raimond.winslow@gmail.com

Raimond L. Winslow, Ph.D., is the Director of the Institute for Computational Medicine, and is the Raj and Neera Singh Professor of Biomedical Engineering at Johns Hopkins University. He is Principle Investigator of the

CardioVascular Research Grid Project. His research interests are computational modeling of the electrical function of the heart in health and disease and cardiovascular informatics.

Marsha Young, J.D.

young_marsha@bah.com

Marsha Young, J.D., has over 34 years of Federal Government experience in the Social Security Administration and the Centers for Medicare and Medicaid Services. She has 5 years of consulting experience primarily supporting the National Institutes of Health. Ms. Young's executive and professional experience has focused on Information Technology management. Before retiring from the Federal government Senior Executive Service, she served as SSA's first full time CIO and established that office and function for the Social Security Administration including the governance structures and procedures for the IT budgeting and portfolio management process. She managed one of the President's initial OMB e-government initiatives that focused on automating access to state-held birth and death records in support of the benefit claims process. Ms. Young serves as a subject matter expert for the cancer Biomedical Informatics Grid® program in the National Institutes of Health, National Cancer Institute. She works with NCI and participating cancer research centers as well as others involved with health data exchanges to address legal and regulatory issues related to patient health data privacy as well as intellectual property and proprietary interests for organizations involved in the exchange of sensitive health data. Ms. Young also provides policy expertise in identity management and secure access control issues for health data exchanges. Ms. Young is licensed to practice law by the State of Maryland, is a member of the American Bar Association, the Maryland Bar Association, the American Health Law Association and the Phi Delta Phi International Legal Fraternity.

Howard Zaren, M.D.

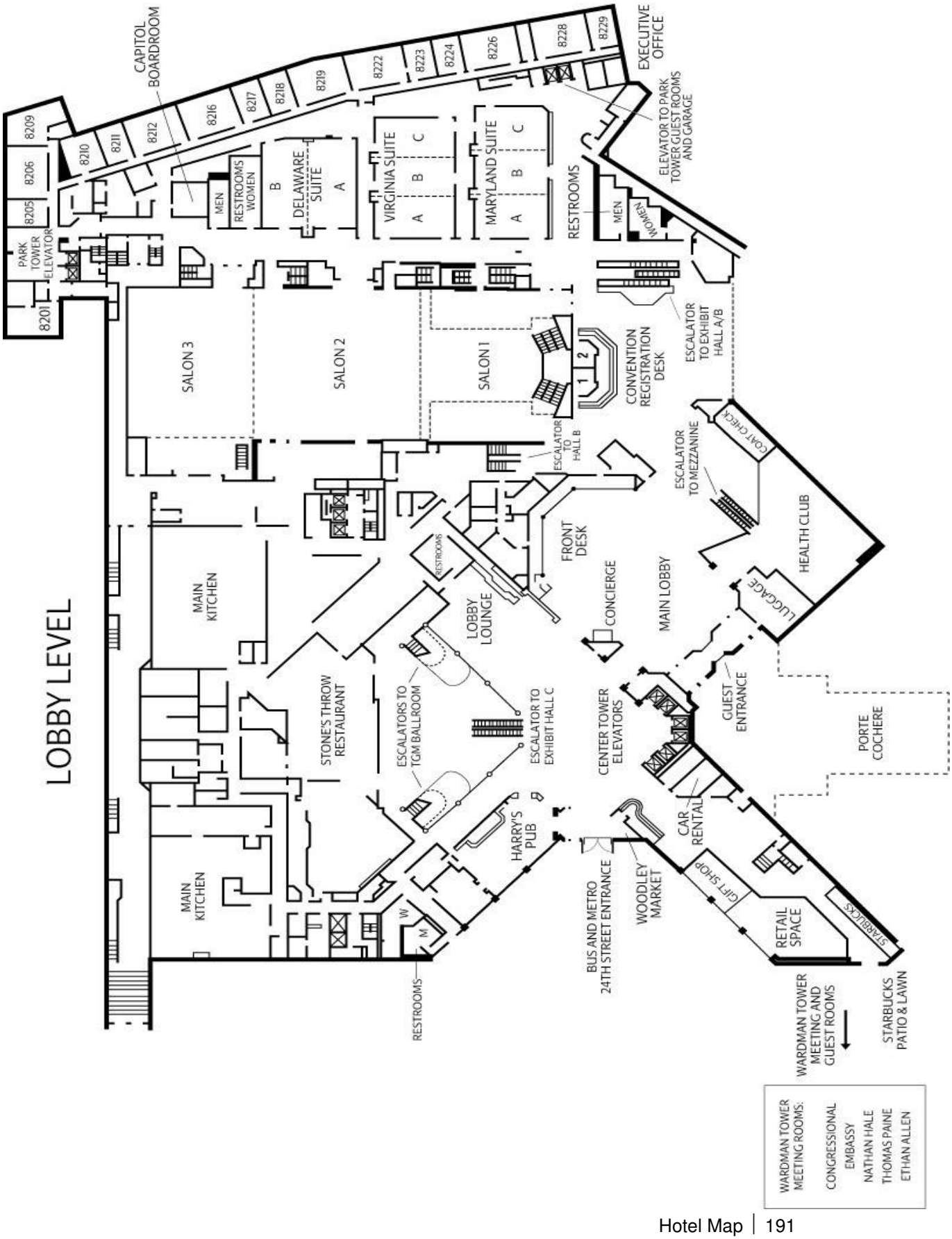
Howard Zaren, M.D., received a bachelor's degree from Long Island University and a doctor of medicine degree from the University of Manitoba Medical School. His post graduate training included a general surgery residency at Pennsylvania Hospital. He completed a fellowship in surgical oncology and gastrointestinal endoscopy at the University of Texas System Cancer Center, M.D. Anderson Hospital & Tumor Institute. After completing his fellowship training he then became a Clinical Assistant Professor of Surgery at the University School of Medicine in Philadelphia and was a private practicing surgical oncologist at Pennsylvania Hospital. He then went on to the Medical College of Pennsylvania where he was the Interim Chair of the Department of Surgery and Tenured Professor of Surgery. In 1990 he was recruited as Chair of Surgery and Cancer Center Director of Mercy Hospital of Pittsburgh and Clinical Professor of Surgery at the University of Pittsburgh. In 1999, he was recruited to be the Chair of the Department of Surgery at the John H. Stroger Hospital of Cook County, Chicago. Soon after he was appointed Tenured Professor of Surgical Oncology at the University of Illinois and the Principal Investigator of the Minority-Based Community Clinical Oncology Program of Cook County. Since 1988, he has published 57 refereed articles, 89 abstracts and 11 book chapters, and has presented at regional, national and international venues. Dr. Howard Zaren became Medical Director of the Nancy N. and J. C. Lewis Cancer & Research Pavilion in Savannah, Georgia in October 2008. He also serves as the Principal Investigator for the National Cancer Institute's National Community Cancer Center's Pilot Program (NCCCP) and is a Georgia Cancer Coalition Distinguished Cancer Scholar.

Jinghui Zhang, Ph.D.

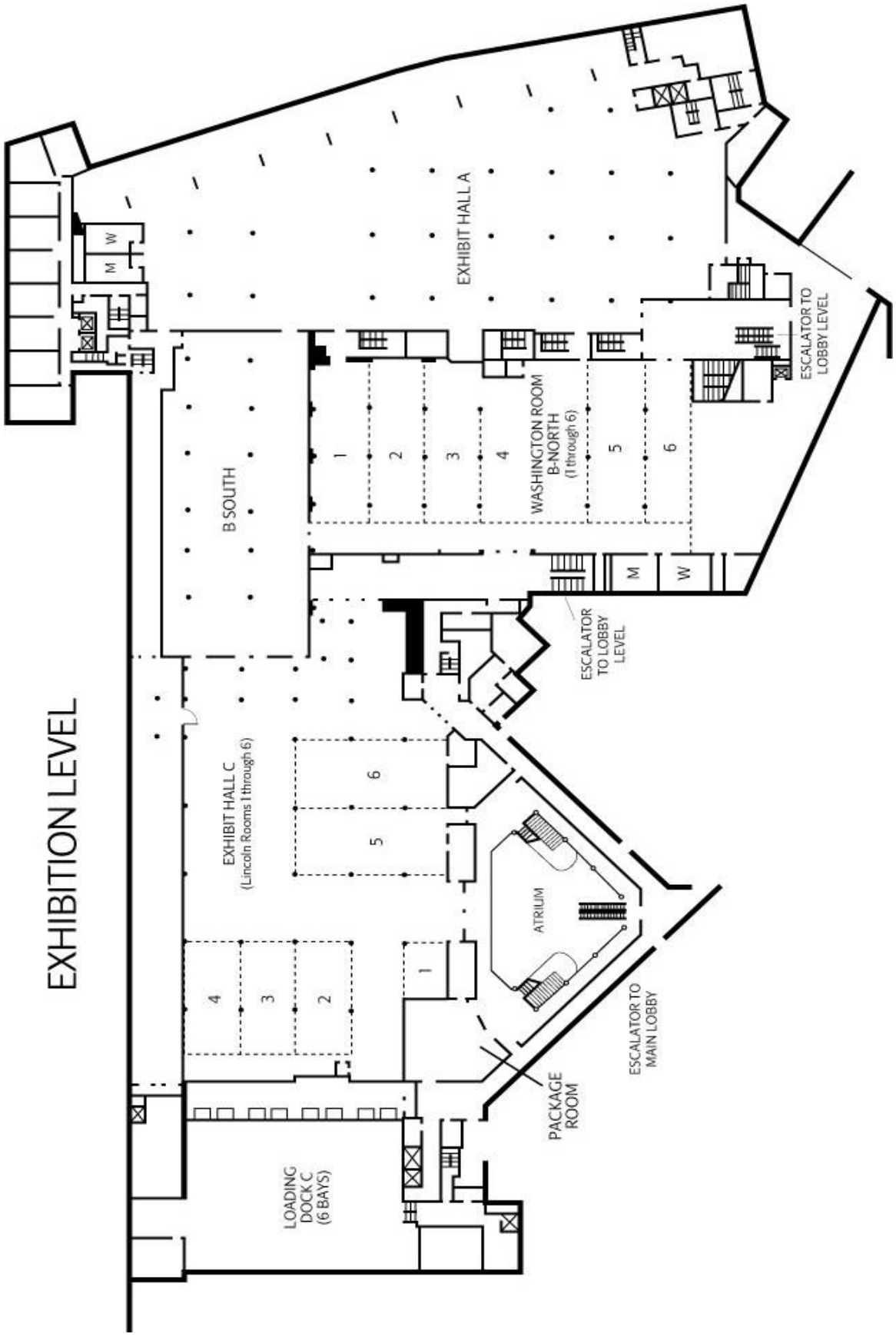
Jinghui.Zhang@STJUDE.ORG

Jinghui Zhang, Ph.D., is a computational biologist interested in understanding the effect of genetic alterations on cancer initiation, progression and prognosis through integrative analysis of large-scale, multi-dimensional genomic data. Her research interest has been in the development of highly accurate and sensitive computational methods for analyzing large-scale genomic data, especially in the area of detecting and analyzing genetic variations and somatic mutations using traditional Sanger sequencing or the next-generation sequencing technology. Building a comprehensive error model for large-scale genomic analysis and developing cross-platform data integrity analysis tools have been her primary research interest. Development of visualization tools for viewing integrative genomic data is another key interest as these tools are indispensable for gaining understanding of the experimental data and for reviewing the accuracy of the computational results obtained from the astronomical volume of the genomic data generated with the current technology. By focusing on understanding genomic data error model, her group has developed computational tools for analyzing genetic variations and somatic mutations. These tools have been used by the two major cancer genomic initiatives of the National Cancer Institute: Cancer Genome Atlas Project (TCGA) and the Therapeutically Applicable Research to Generate Effective Treatment (TARGET). She has also actively participated in the analysis of these initiatives and is the first to discover recurrent *NF1* mutations in GBM, near 100% *TP53* mutations in ovarian cancer for the TCGA project, recurrent *JAK* mutations in pediatric ALL in the COG/TARGET project. Understanding the genetic origin of pediatric cancers is currently the main focus of her research as she is primarily responsible for analyzing the next-generation sequencing data generated from the St. Jude Children's Research Hospital - Washington University Pediatric Cancer Genome Project. Developing novel approaches that integrate whole-genome sequencing data with RNA sequencing, copy number, gene expression is an ongoing effort. Ongoing collaborative research with colleagues in structural biology and statistics will allow us to gain further understanding of functional impact of recurrent somatic alterations on pediatric cancers.

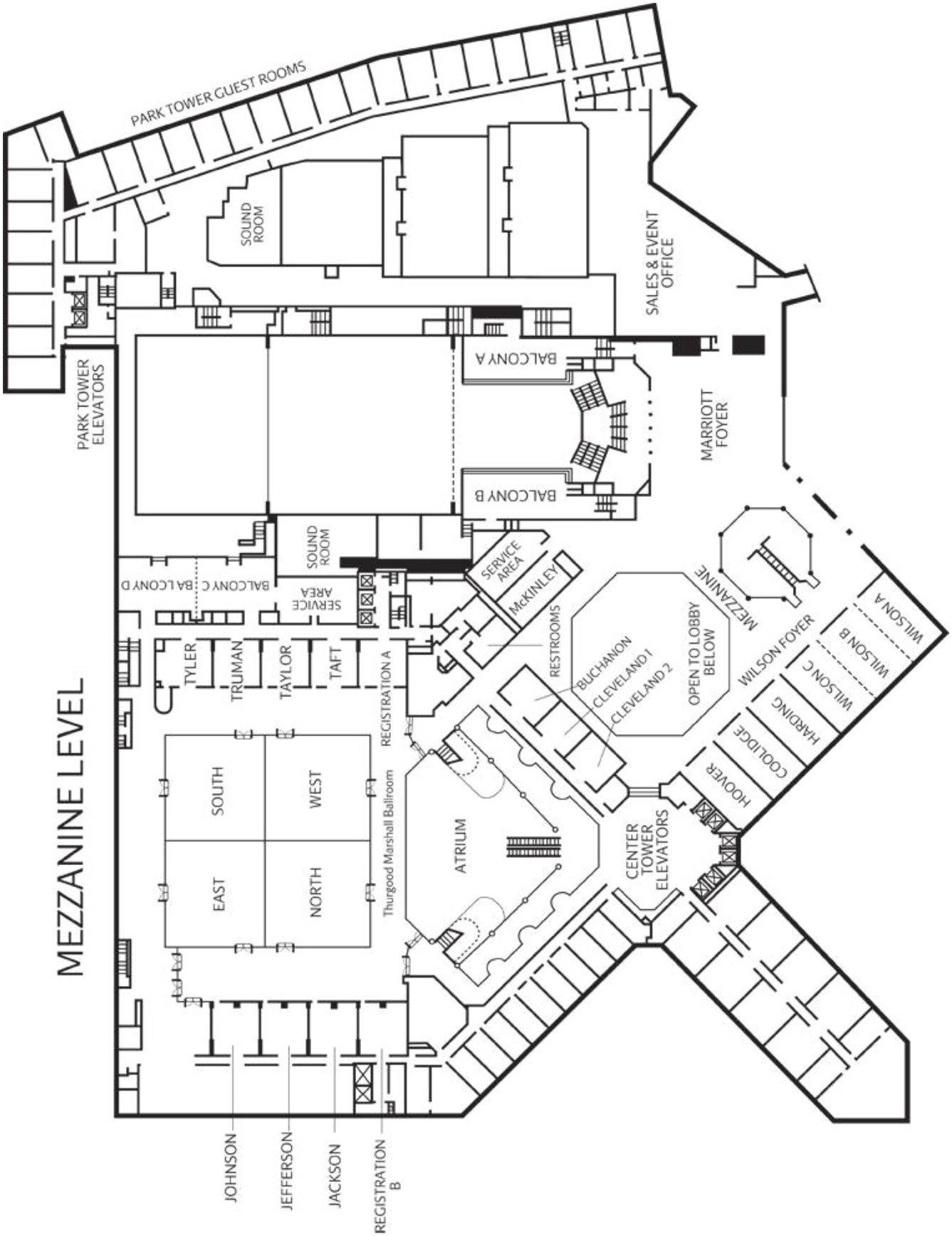
Hotel Map – Lobby Level



Hotel Map – Exhibition Level



Hotel Map – Mezzanine Level



Map of the Washington, DC Metro System

NATIONAL® CANCER INSTITUTE