Putting Data to Work—Creating Value-Added Functionality and Applications Through Mashups

Dominic James DiFranzo, *Graduate Student, Computer Science, Rensselaer Polytechnic Institute*

George Komatsoulis, *Deputy Director of Center for Biomedical Informatics and Information Technology, NCI, NIH*

Ryan McKeel, *Digital Assets Applications Engineer, U.S. National Renewable Energy Laboratory*

Susan Turnbull, *Senior Program Advisor, Office of Citizen Services and Innovative Technologies (Moderator)*
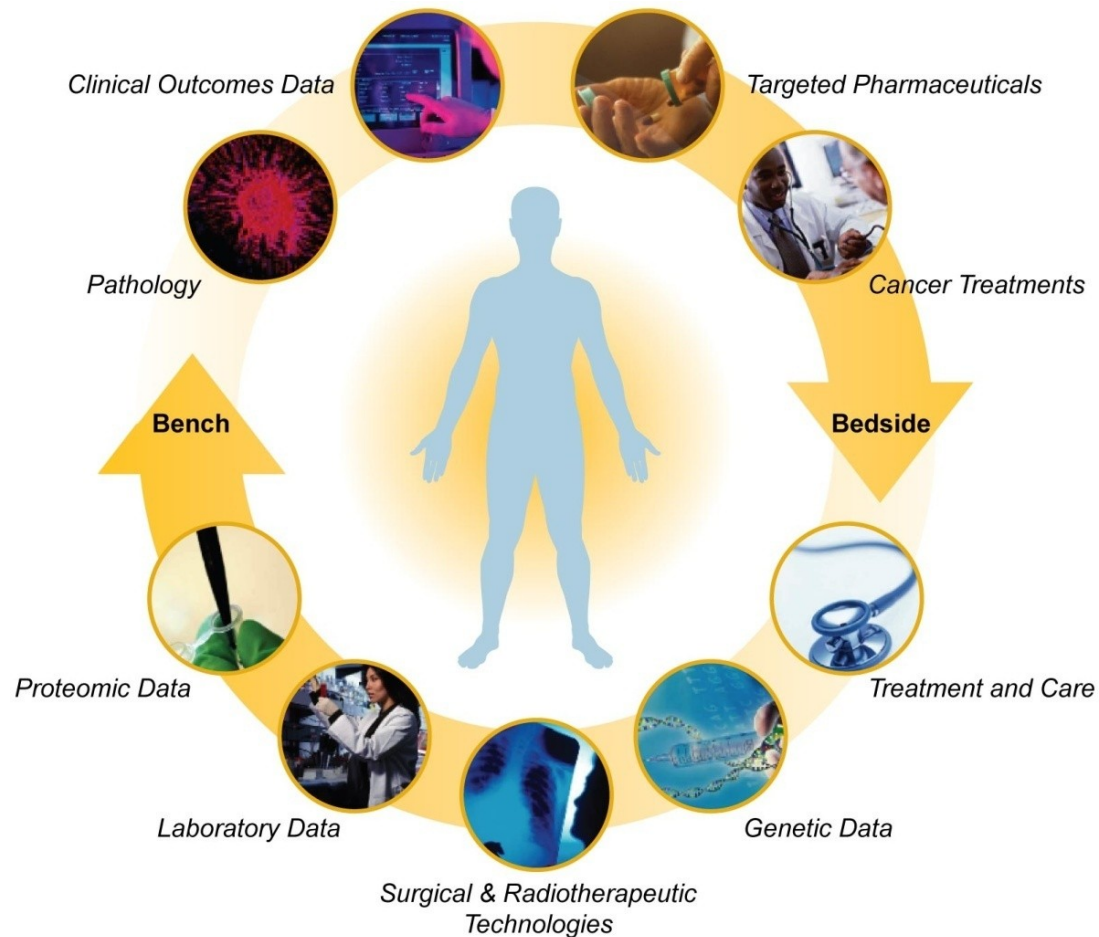
The cancer Biomedical Informatics Grid

International Open Government Data Conference

November 15, 2010

George A. Komatsoulis, Ph.D.
*Deputy Director, Center for Bioinformatics and Information Technology, NCI*

21st Century Biomedical Research and Care

- Personalized, Predictive, Preemptive, Participatory……

- Unifies discovery, clinical research, and clinical care, (bench-bedside-bed) into a seamless continuum

- Results in improved clinical outcomes

- Accelerates the time from discovery to patient benefit

- Enables a health care system, not a disparate "sector"

- Empowers consumers in managing health over a lifetime

Clinical Outcomes Data
Targeted Pharmaceuticals
Pathology
Cancer Treatments
Bench
Bedside
Proteomic Data
Treatment and Care
Laboratory Data
Genetic Data
Surgical & Radiotherapeutic Technologies

Molecular Medicine and Childhood ALL

Increased 5-year relative survival rate among children 10-14 years old diagnosed with ALL
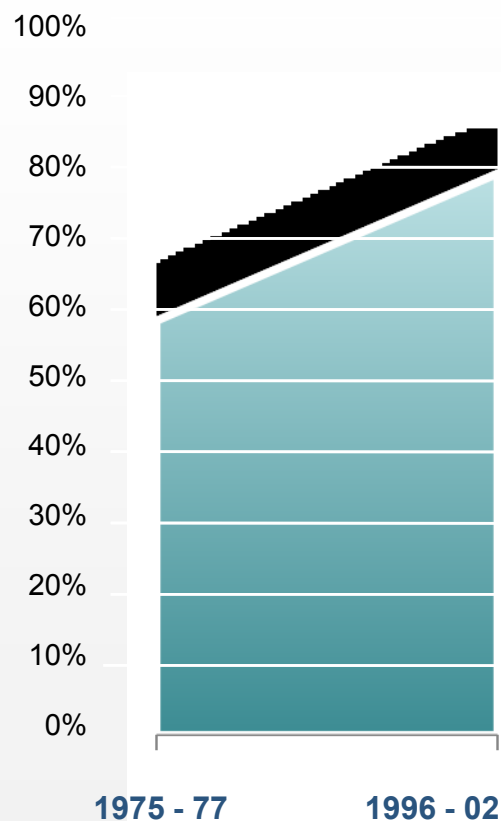
Childhood cancer treatment now blends care delivery and clinical research. Most children with cancer are enrolled in clinical trials. Using this approach:

- Researchers and practitioners are able to correlate experimental laboratory data with clinical data such as treatment, history, pathology and patient outcomes

- Clinical data are used to continuously evaluate outcomes

- Researchers can develop and refine evidence-based strategies at an individualized level

- Care providers improve quality by adherence to care standards

Linking the NCI Supported Cancer Community

■ - NCI Desi

▲ - Community Oncology Practices

caBIG®: Creating a Worldwide Web of Cancer Research

caBIG® **is 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.**

caBIG® Core Principles

- **Open Access** – caBIG® is open to all, enabling wide-spread access to tools, data, and infrastructure

- **Open Development** – Planning, testing, validation, and deployment of caBIG® tools and infrastructure are open to the entire research community

- **Open Source** – The underlying software code of caBIG® tools is available for use and modification

- **Federation** – Resources can be controlled locally, or integrated across multiple sites

cancer Biomedical Informatics Grid® | caBIG®

caBIG® Strategy

- **Community**
  - Establish an **open community** of participants from the spectrum of disciplines, geographies, types of institutions, etc.
  - Facilitate the work of others who are building capabilities
  - Adopt a **"federated"** model to allow local control of sharing and partnerships and to support individual labs and institutions
- **Content**
  - Facilitate access to rich primary data
  - Leverage existing academic and commercial software, wherever possible, to avoid unnecessary time and expense
  - Invest primarily in **open source** tools that the community does not have
- **Connectivity**
  - Recognize legacy IT systems to avoid "rip and replace" costs
  - Wherever feasible, make disparate applications compatible for "plug-and-play" compatibility and data-sharing through **standards-based interoperable infrastructure**

caBIG® Operational Approach

Domain-level

Strategic-level

Clinical Trials Management Systems Workspace (CTMS)

Integrative Cancer Research Workspace (ICR)

In Vivo **Imaging Workspace (IMAG)**

Tissue Banks & Pathology Tools Workspace (TBPT)

Strategic Planning Workspace (SP)

Training Workspace (D&T)

Data Sharing & Intellectual Capital Workspace (DSIC)

Cross-cutting

caBIG® Vocabularies and Common Data Elements Workspace (VCDE)

caBIG® Architecture Workspace (ARCH)

cancer Biomedical Informatics Grid® | caBIG™

Interoperable Applications to Support Biomedical Research

Clinical Research

- Track clinical trial registrations
- Facilitate automatic capture of clinical laboratory data
- Manage reports describing adverse events during clinical trials

Imaging

- Utilize the National Biomedical Imaging Archive repository for medical images including CAT scans and MRIs
- Visualize images using DICOM-compliant tools
- Annotated Images with distributed tools

Molecular Biology

- Combine proteomics, gene expression, and other basic research data
- Submit and annotate microarray data
- Integrate microarray data from multiple manufacturers and permit analysis and visualization of data

Pathology

- Access a library of well characterized, clinically annotated biospecimens
- Use tools to keep an inventory of a user's own samples
- Track the storage, distribution, and quality assurance of specimens

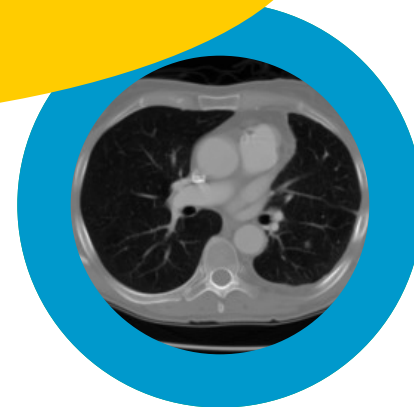# Supporting Individual and Institutional Needs

Analytical Tools
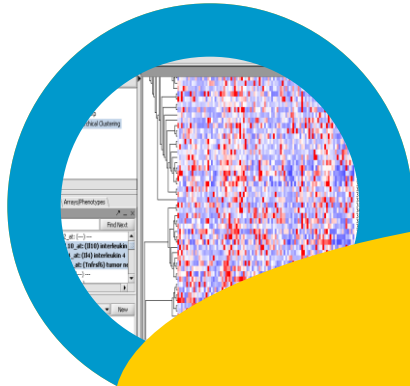
Molecular Analysis

Bios... **...als**

Common Data Elements

Vocabularies & Ontologies

Information Models

caBIG Integrates Organizational IT Infrastructure

- **caGrid Technology can be used locally to integrate Biomedical IT elements within an institution, examples include:**

  - **Ohio State University**: Using caGrid to create the TRIAD system to support clinical research integration

  - **Washington University**: Integrating clinical trials, EHR and biospecimen resources

  - **University of Alabama, Birmingham**: Using caGrid as their internal integration technology

  - **University of Arkansas Medical School**: Integrated clinical trials information system leveraging caGrid

  - **Aga Khan University**, Karachi, Pakistan

cancer Biomedical Informatics Grid® | caBIG

NCI Hosted caGrid Deployment
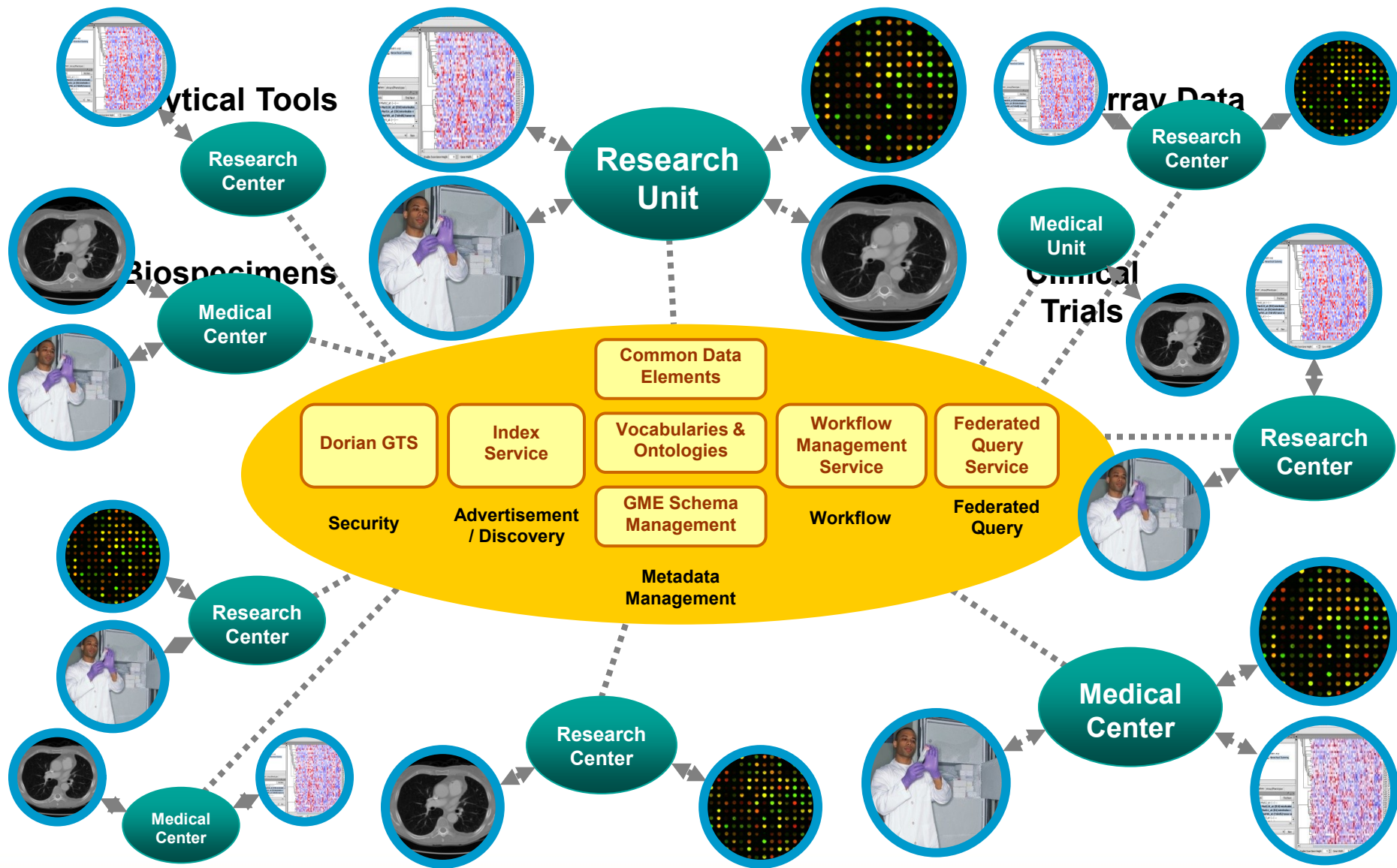
caGrid in use outside of caBIG®

- **A variety of organizations are adopting or adapting caGrid to build interoperability Infrastructures. Examples include:**

  - **National Cancer Research Institute (NCRI) UK:** Using caGrid as the basis for their ONIX infrastructure

  - **NHLBI CardioVascular Research Grid (CVRG):** NCI caGrid and CVRG have begun cross-indexing services to allow access to capabilities available on both Grids

  - **Clinical Translational Science Awards (CTSA):** Ohio State University implementing caGrid for CTSA awardees

  - **Centre for the Development of Advanced Computing (CDAC), Ministry of Information Technology, Government of India:** Implementing caGrid to bring Indian research organizations into the community of cancer research.

- **caGrid Modular Infrastructure allows the creation of a "Grid of Grids"**

caBIG® at a glance

Community

- 2,300+ participants from more than 700 institutions
  - 56 NCI-designated Centers
  - 30 Community Centers
- 1000+ registrants for the 2010 caBIG® Annual Meeting
- 19 licensed Support Service Providers to sustain the biomedical community as they deploy caBIG® tools and technology
- 15 countries using caBIG® tools and technology to facilitate

Connectivity

- 78 applications supporting full continuum of biomedical research
- 149 "nodes" connected to National Grid via caGrid

Content

- 2.17 million biospecimens available through caGrid
- 4.76 million images stored in the National Biomedical Imaging Archive
- 39,952 microarray experiments available for research use on caGrid

Expanding Across Sectors and Internationally

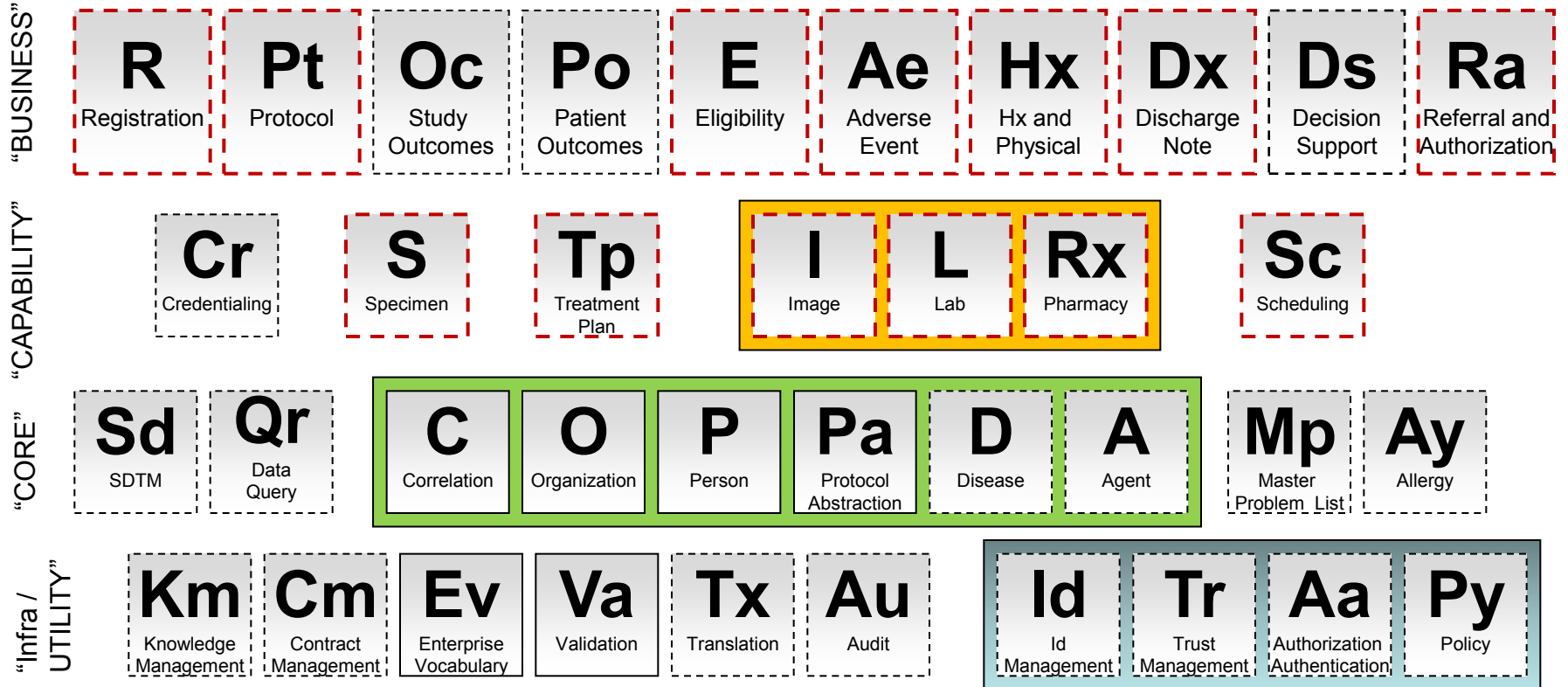**15 countries engaged with and/or using caBIG® tools and technologies, including:**

- United Kingdom
- Latin America
- India
- China
- Mexico, Brazil, Uruguay, Argentina, Chile
- Czech Republic
- The Netherlands
- Germany
- Finland
- Jordan
- Pakistan
- Australia
- New Zealand

- **United Kingdom:** NCI collaboration with National Cancer Research Initiative (NCRI) focused on the use of caGrid technology to connect researchers and enable exchange of research data.

- **Latin America:** Latin American Cancer Pilot Program began patient enrollment for two clinical trials on molecularly characterized stage II and III breast cancer patients, using a broad suite of caBIG® tools to ensure that researchers can compare data across partner sites.

- **China:** Duke University Comprehensive Cancer Center and Beijing University Cancer Hospital launched a collaboration using caBIG® capabilities to conduct first clinical trials in China where all patients are registered electronically.

India: NCI engaged with the All India Institute of Medical Sciences, the Center for Development of Advanced Computing, and the Tata Memorial Hospital of Mumbai over the use of grid computing for managing clinical trials data.

cancer Biomedical Informatics Grid® | caBIG

Enterprise Services for Research and Care

"BUSINESS"

R Registration	**Pt** Protocol	**Oc** Study Outcomes	**Po** Patient Outcomes	**E** Eligibility	**Ae** Adverse Event	**Hx** Hx and Physical	**Dx** Discharge Note	**Ds** Decision Support	**Ra** Referral and Authorization

"CAPABILITY"

Cr Credentialing	**S** Specimen	**Tp** Treatment Plan	**I** Image	**L** Lab	**Rx** Pharmacy	**Sc** Scheduling

"CORE"

Sd SDTM	**Qr** Data Query	**C** Correlation	**O** Organization	**P** Person	**Pa** Protocol Abstraction	**D** Disease	**A** Agent	**Mp** Master Problem List	**Ay** Allergy

"Infra / UTILITY"

Km Knowledge Management	**Cm** Contract Management	**Ev** Enterprise Vocabulary	**Va** Validation	**Tx** Translation	**Au** Audit	**Id** Id Management	**Tr** Trust Management	**Aa** Authorization Authentication	**Py** Policy

cancer Biomedical Informatics Grid® | caBIG

Next Generation: caBIG® 2.0

caBIG® capabilities are evolving to support the rapidly changing needs of Cancer Centers as they move to EHRs and genomically-guided medicine:
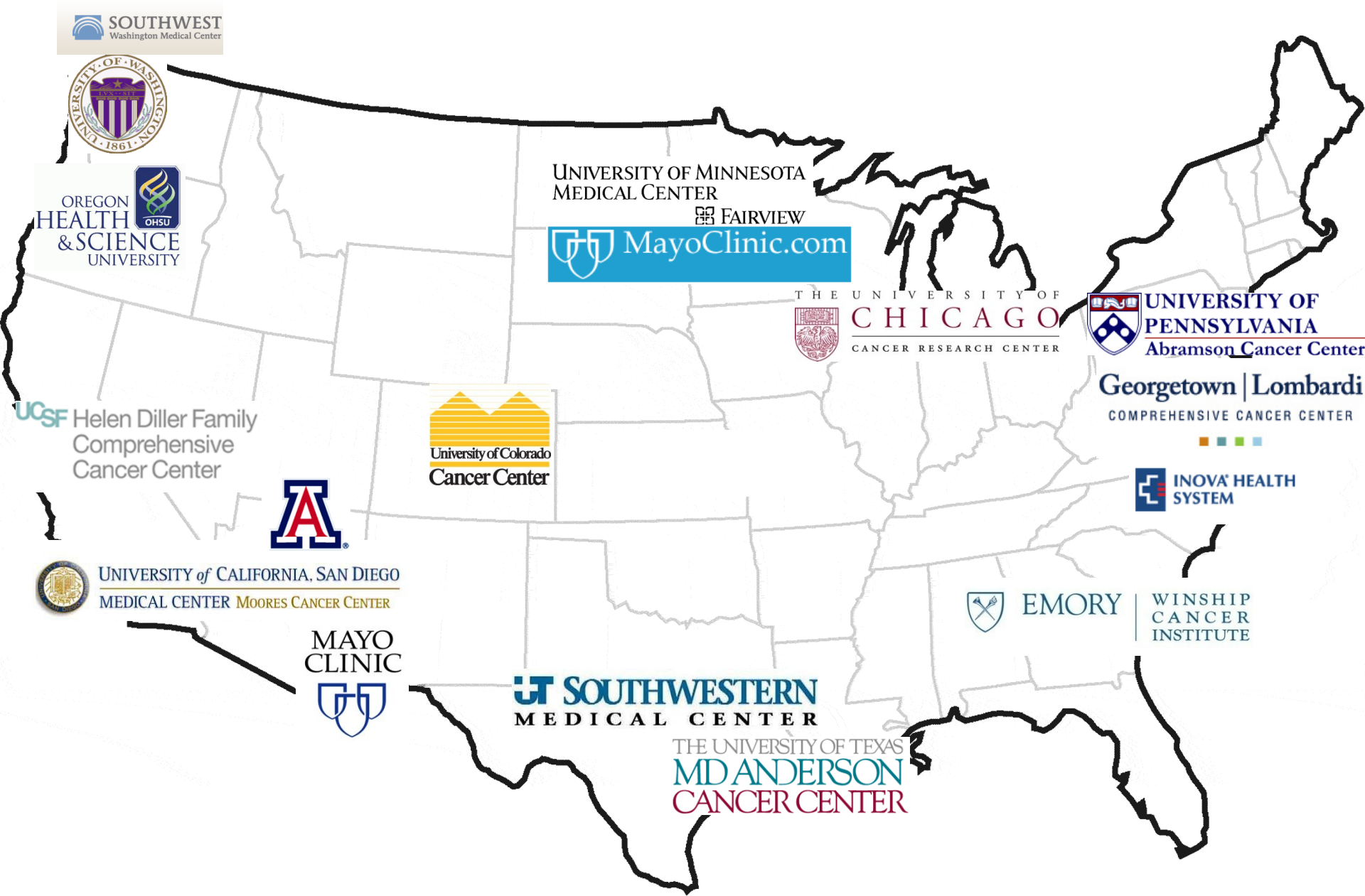
- Collaborate with ASCO on caEHRs
- Capitalize on tech advances: cloud computing, Service-Oriented Architecture, mobile devices
- Lower barriers: make it easier and easier for all stakeholders to use caBIG® research capabilities
- Leverage increasingly mature collection of publicly-available open source infrastructure

The I-SPY TRIAL (Investigation of Serial studies to Predict Your Therapeutic Response with Imaging And moLecular analysis):

A national study to leverage biomarkers
in predicting response to combinatorial therapy for
women with Stage 3 breast cancer.

(PI Laura Esserman, UCSF)

Projected I-SPY 2 study sites
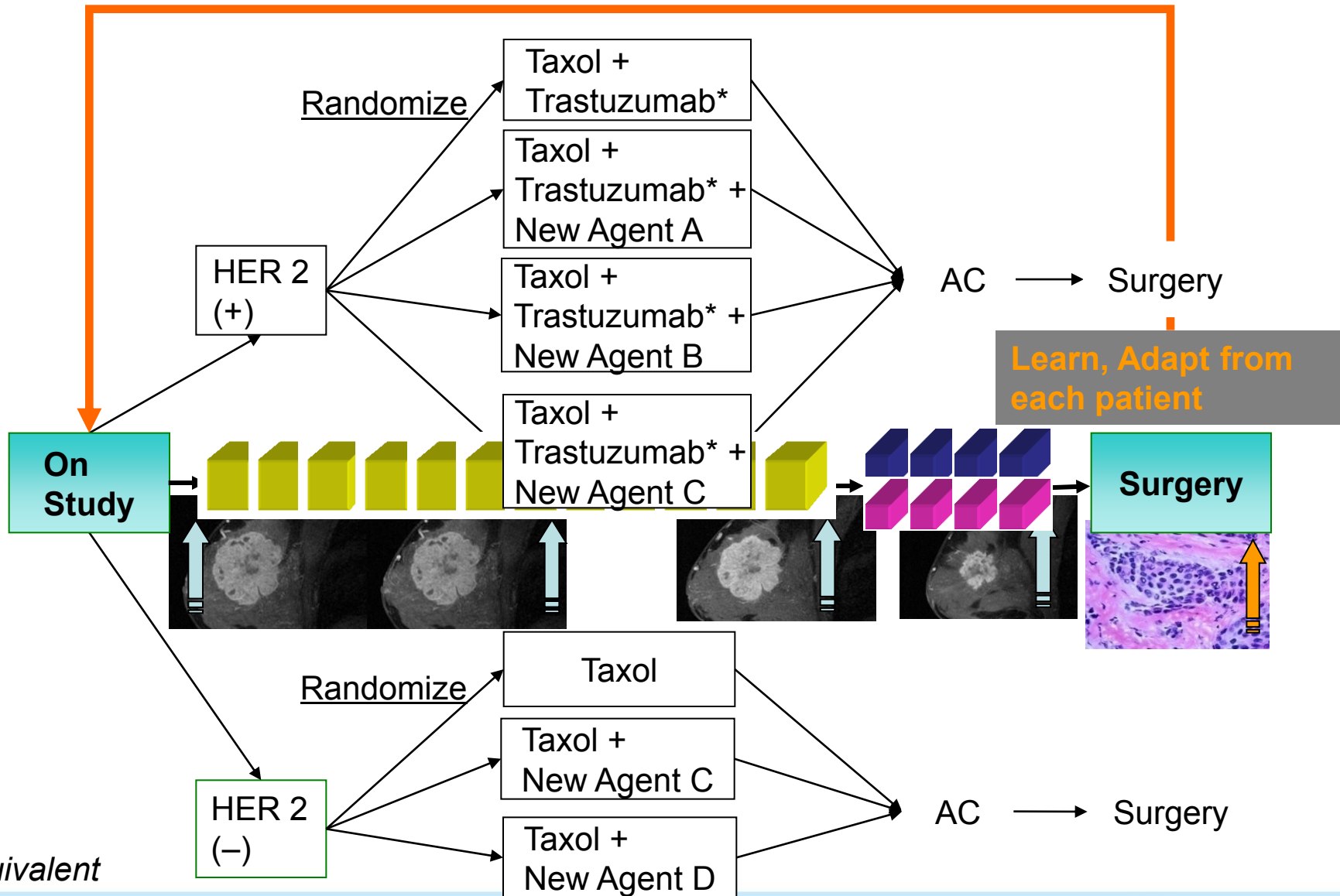
I-SPY Adaptive Trial Outline

Taxol +/−New Drug
(12 weekly cycles)

AC
(4 cycles)

On Study

Surgery

Biopsy
Blood

MRI

MRI
Biopsy
Blood

MRI
Blood

MRI
Blood

Tissue
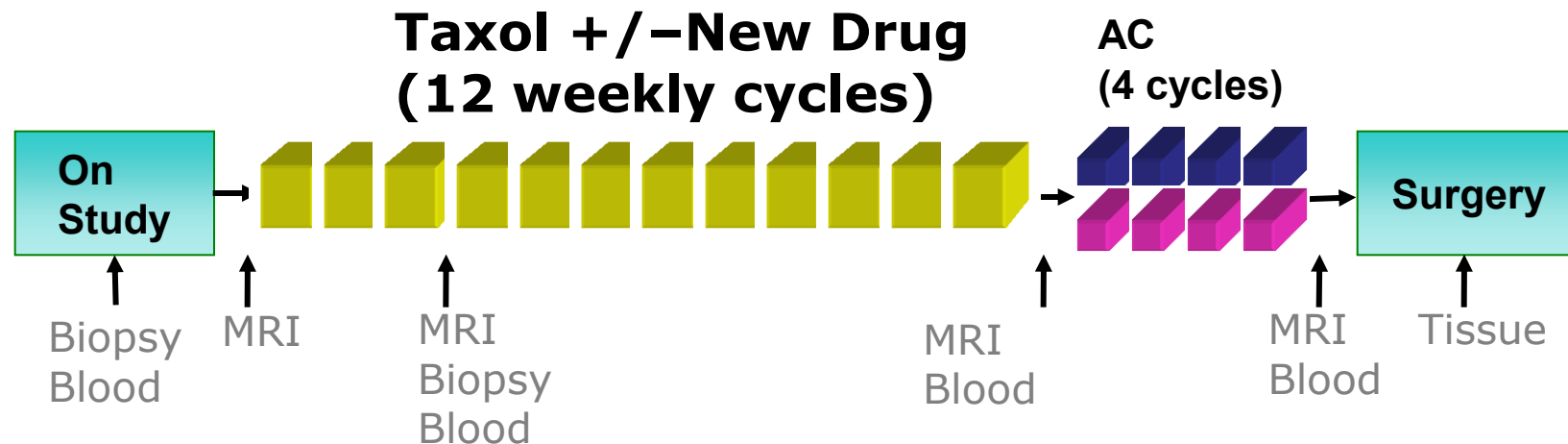
Accrual: Anticipate 800 patients over 3−4 years

Enroll ~20 patients per month

Participating Sites: 15−20 across US and Canada
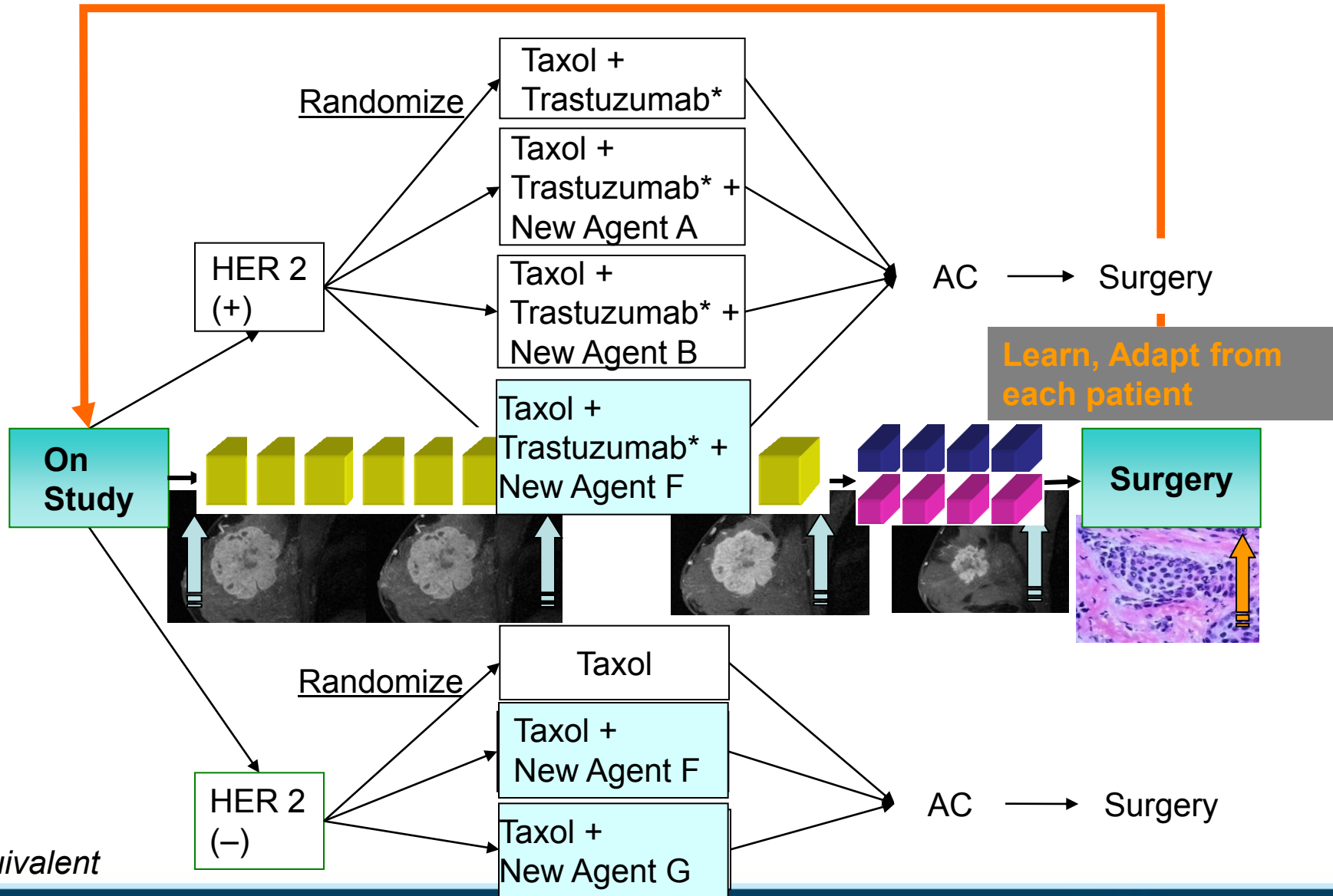
I-SPY Adaptive Trial:
Introduce several new agents for a given profile

*Or Equivalent

I-SPY Adaptive Trial:
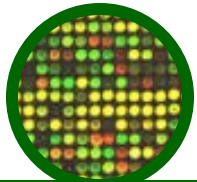
Introduce several new agents for a given profile

On Study

HER 2 (+)

Randomize

- Taxol + Trastuzumab*
- Taxol + Trastuzumab* + New Agent A
- Taxol + Trastuzumab* + New Agent B
- Taxol + Trastuzumab* + New Agent F

AC → Surgery

Learn, Adapt from each patient

Surgery

HER 2 (−)

Randomize

- Taxol
- Taxol + New Agent F
- Taxol + New Agent G

AC → Surgery

Or Equivalent

cancer Biomedical Informatics Grid® | caBIG

I-SPY TRIAL IT Infrastructure
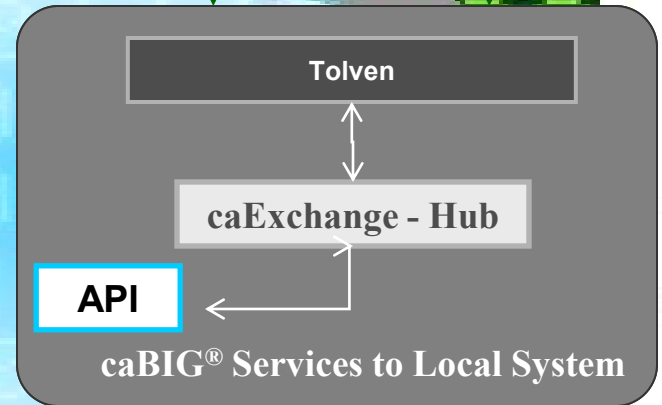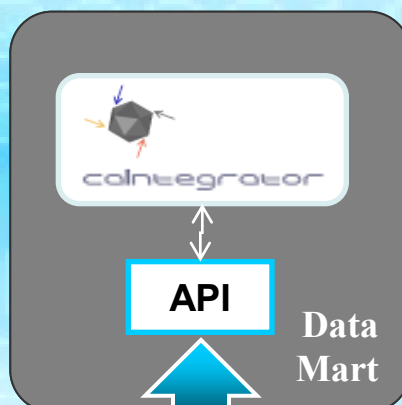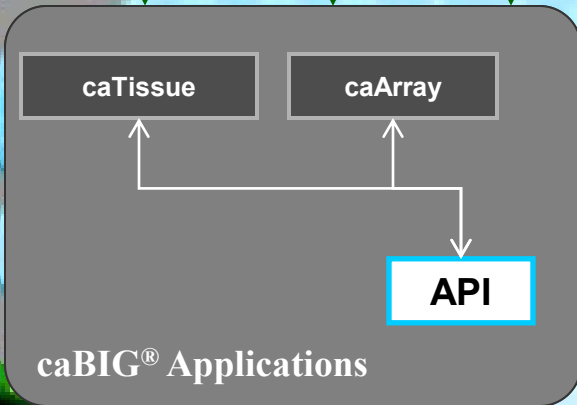
For more information, please visit:

http://caBIG.cancer.gov

http://www.bighealthconsortium.org

Thank You