Community – Oriented Informatics:

The caBIG® Program

Sixth SHRP 2 Safety Research Symposium

July 14, 2011

R. Mark Adams, Ph.D.
Principal, Booz Allen Hamilton
caBIG Scientific Program Manager

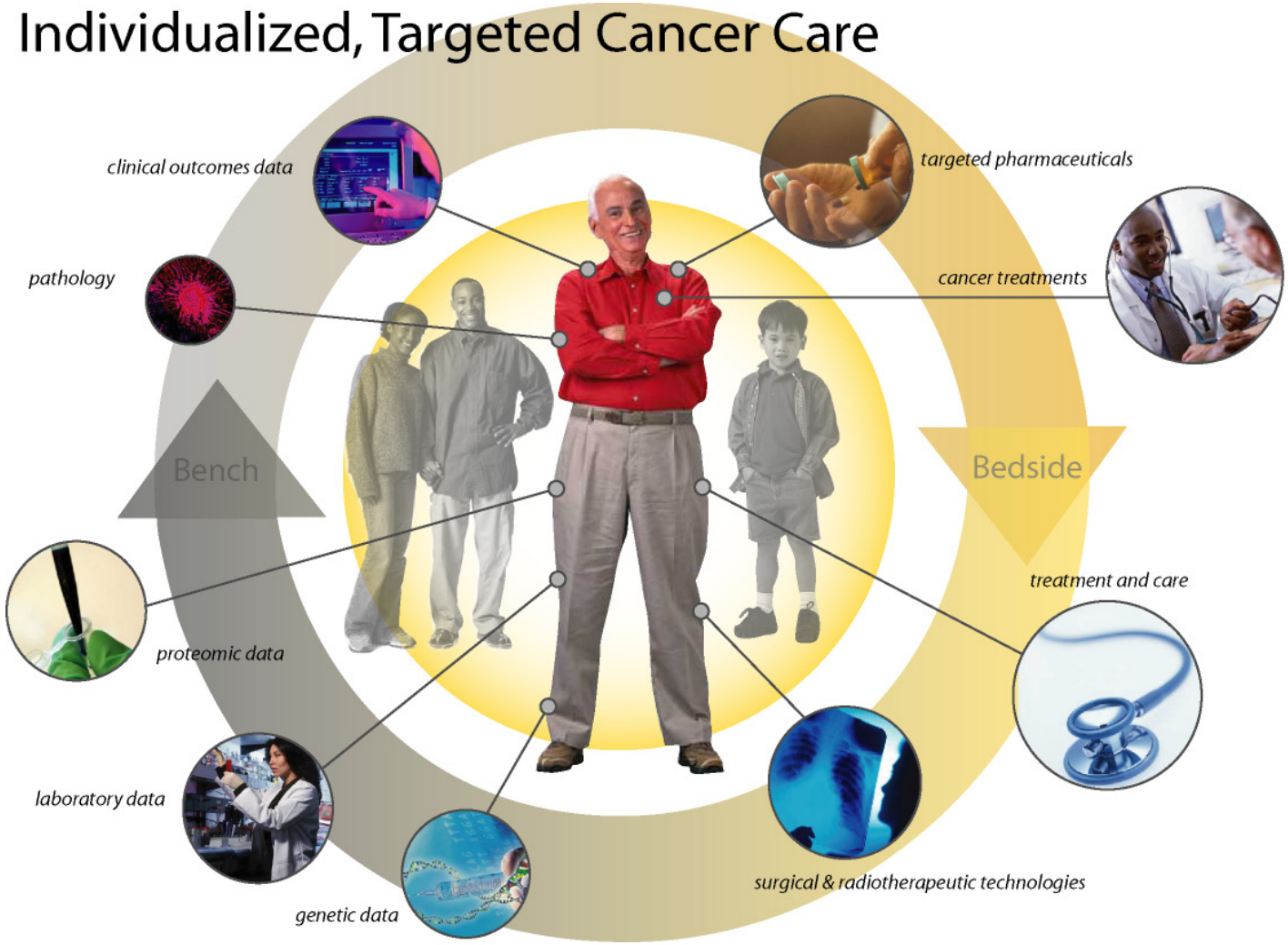
Background and Origins

caBIG®: The Problem We Were Tasked to Solve

- **circa 2002: NCAB charges NCI to assist the Cancer Community with the increasingly unmanageable informatics challenges associated with clinical research, biospecimen management, and molecular analysis**

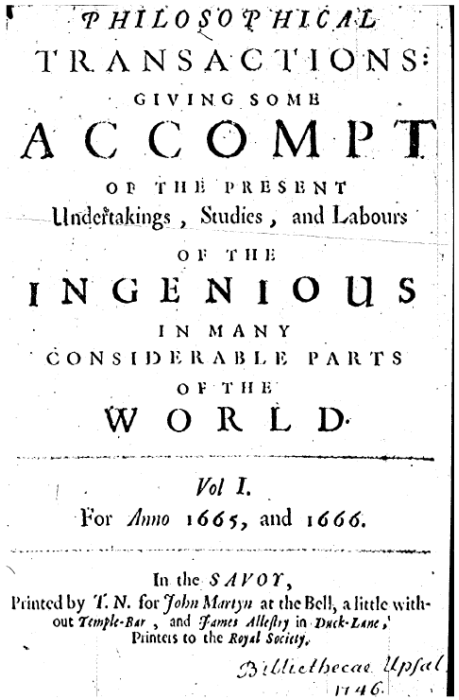
Challenge: provide interoperable resources and capabilities to individuals/institutions/organizations that meet critical needs and address important science questions

Molecular Medicine: Personalized Approaches



Scientific Information Exchange

17th Century



PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.

Vol I.
For Anno 1665, and 1666.

In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and James Allestry in Duck-Lane,
Printers to the Royal Society.

Bibliothecae Upsal
1746.

Royal Society of London
- Oldest learned society (1660)
- Oldest scientific journal (1665)

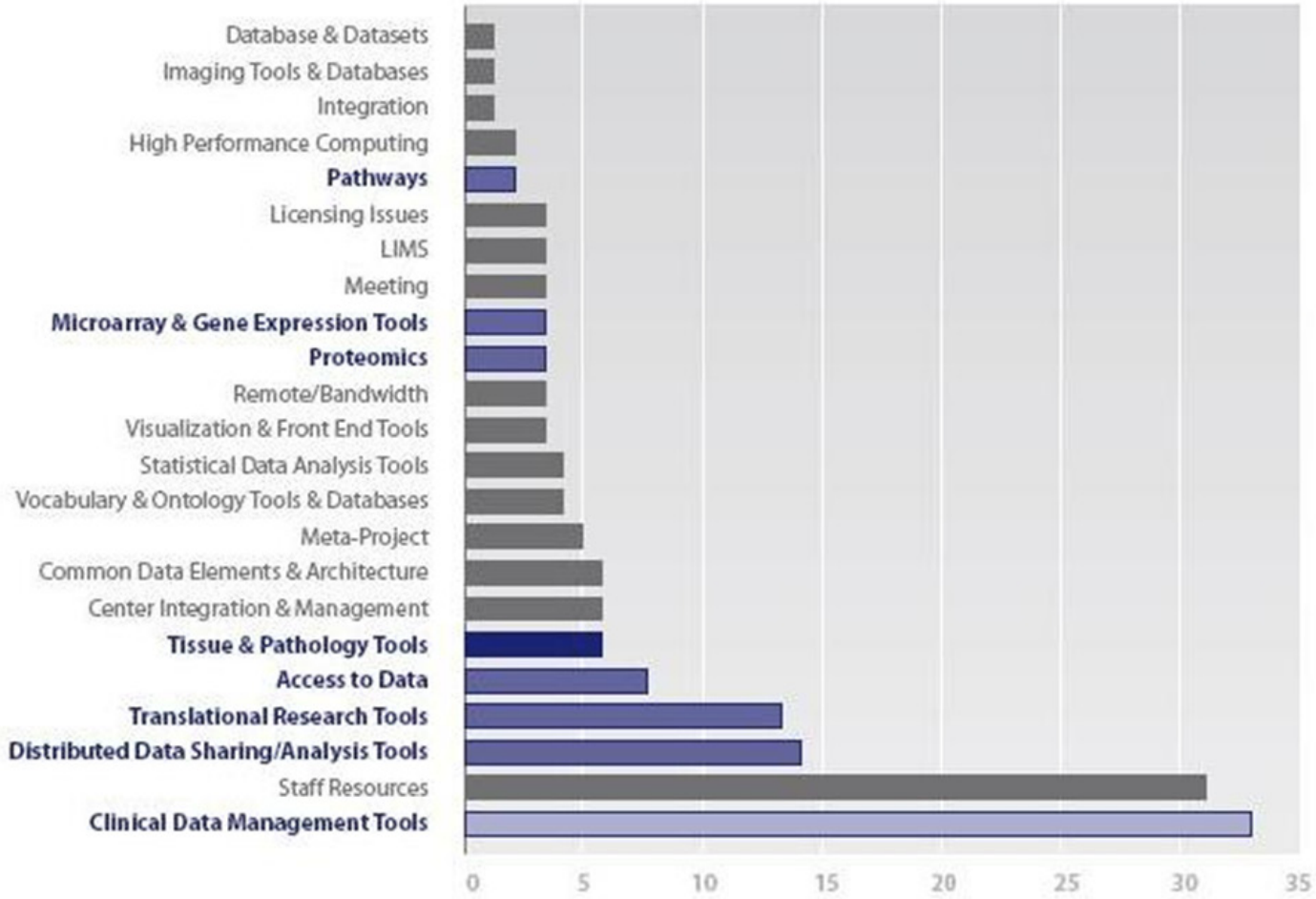
21st Century

Strategic Evolution

"We are now poised for the third big transformation, which will come from behavioral changes, as all participants in the ecosystem – patients, physicians, payers, companies and more – revisit and realign their practices in order to improve health outcomes."

Ernst & Young, Progressions 2011

Input: Cancer Center Needs Determined from Cooperative Development Meetings (2003-04)



Cancer Center involvement identified priority areas for caBIG™

caBIG® Program Structure

caBIG® Program Structure

Domain-level

- Clinical Trials Management Systems Workspace (CTMS)
- Integrative Cancer Research Workspace (ICR)
- *In Vivo* Imaging Workspace (IMAG)
- Tissue Banks & Pathology Tools Workspace (TBPT)

Strategic-level

- 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)

Workspace Roles and Activities

Workspaces have active engagement for and with the community:

- **Bring together, within each domain, funded and volunteer community members engaged in development and adoption**
- **Serve as key operational units of caBIG® and as representatives of the larger community**
- **Open to all who wish to participate**
- **Convened through regularly scheduled teleconferences, webinars, and face-to-face meetings**
- **All products openly and publicly shared through website, forums, wikis and listservs.**

Data Sharing and Intellectual Property

Data Sharing and Intellectual Capital (DSIC) Workspace

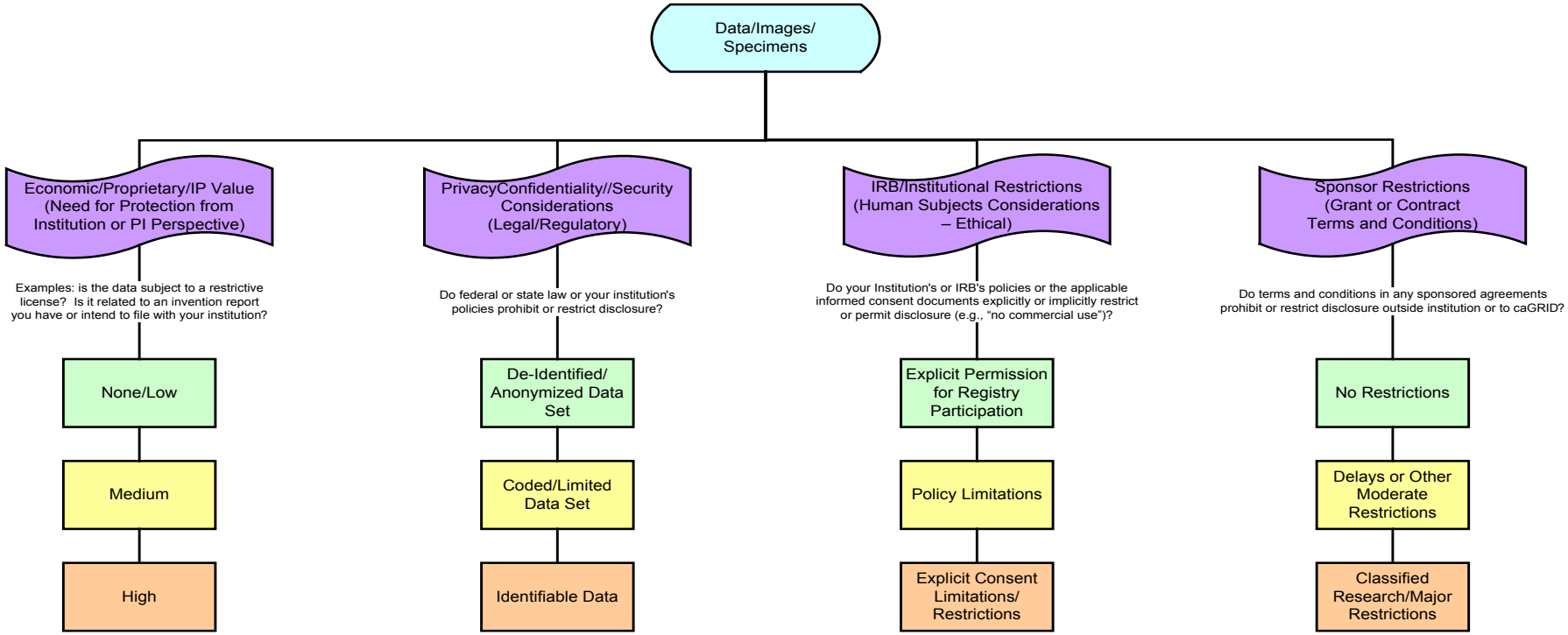
- **Mission of DSIC Workspace**:
  - facilitate data sharing between and among caBIG™ participants by addressing legal, ethical, regulatory, policy, proprietary, and contractual barriers to data exchange

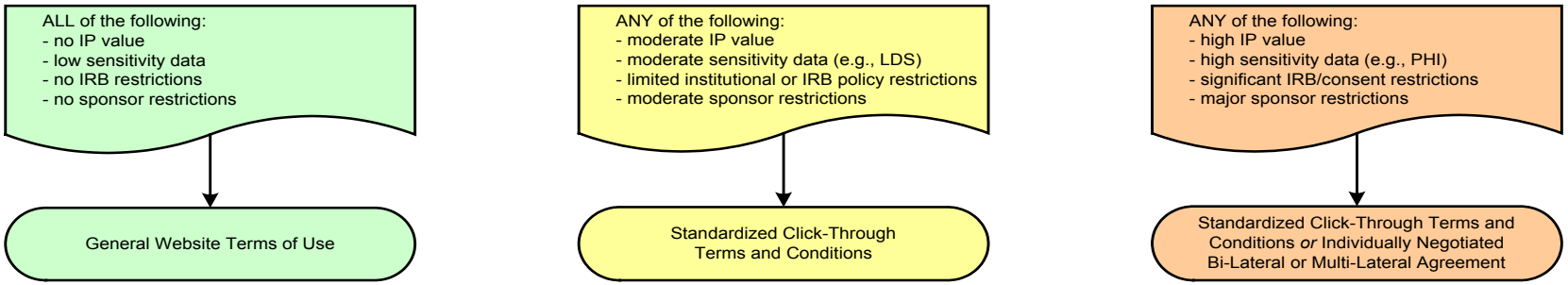
- **Diverse membership:**
  - biomedical researchers, clinicians, technology transfer experts, intellectual property and regulatory attorneys, policy specialists, patient advocates, bioethicists, and bioinformaticists

Data Sharing Challenges for the caBIG® Community

- **Varying obligations under federal and state privacy and security laws/standards**
- **Oversight of human subjects research by ethical review boards (IRBs) -- local requirements vary substantially based on interpretations of applicable requirements**
- **Academic considerations such as the need to secure grants and publish results in peer reviewed literature**
- **Researcher, institutional and sponsor concerns re: protection of intellectual property; industry funding/MTAs often restrict data sharing**
- **Patient safety concerns related to premature access to unvalidated information**
- **Public perceptions regarding privacy, security and confidentiality of electronic health data**

caBIG™ Data Sharing and Security Framework





Case Study

Applying the Framework to TCGA

- **Applying the "Framework" to TCGA**
- **Analysis of potential challenges to data sharing in each of the four identified areas:**
  - Sponsor restrictions
  - Privacy Restrictions
  - IRB/Ethical Restrictions
  - IP restrictions
- **How do we facilitate appropriate exchange of data and specimens to advance scientific knowledge?**

TCGA: Assumptions

- **Data are not subject to any sponsor restrictions.**
- **TCGA publication policies require users to acknowledge TCGA in publications.**
- **IP concerns are otherwise minimal vis. raw, normalized, and segmentation data; IP concerns are significant vis. SNP data.**

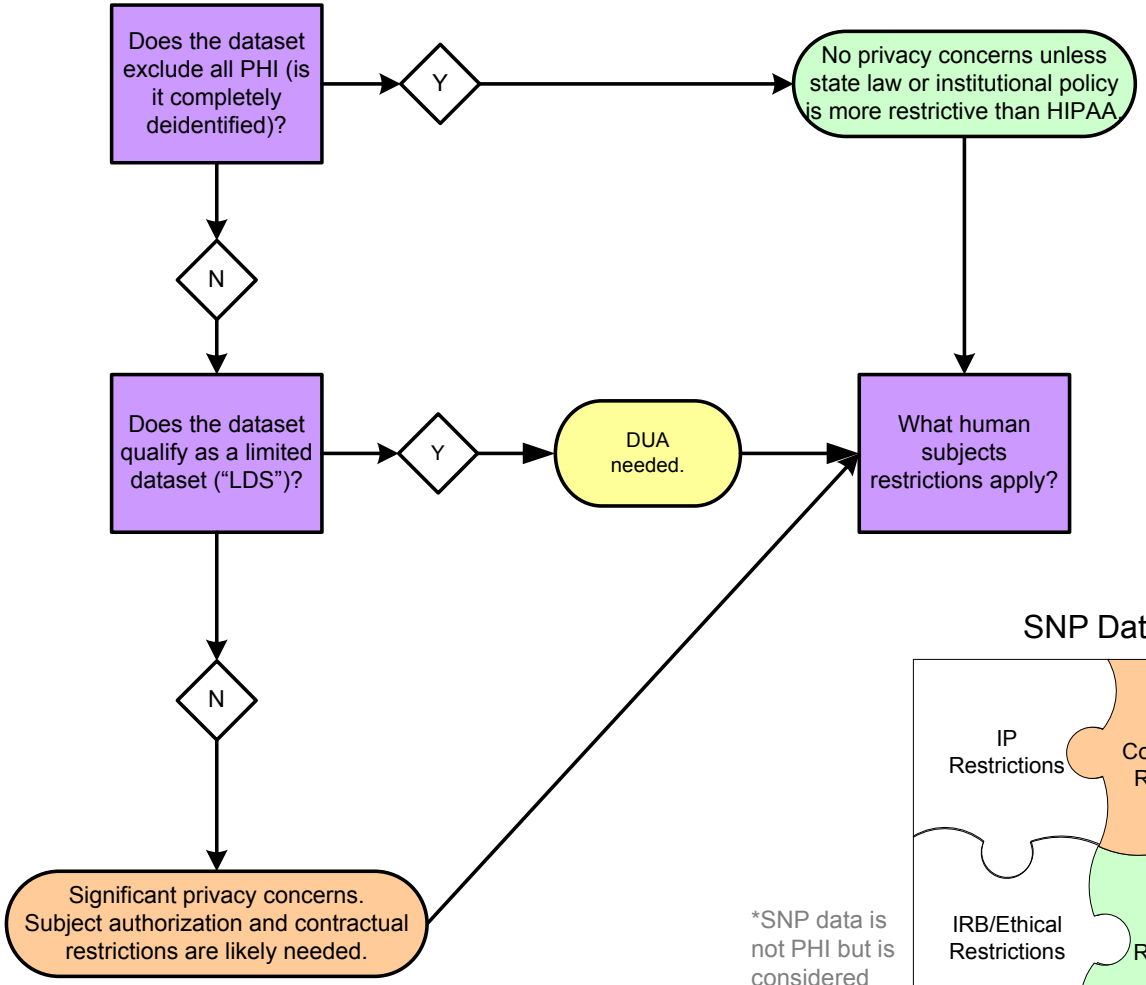
Center/Dataset	Content	Potential for donor ID	Access Policy
BCR complete set	Detailed Phenotype and Outcome data in the following areas (see attached spreadsheet for specific data elements): •Patient •Examination •Surgery •Drugs •Radiation •Sample •TumorPathology •Portion •Slide •Protocol •Analyte •Aliquot	Yes	DAC approval required
BCR minimal clinical dataset	•Clinical Diagnosis •Histologic Diagnosis •Tissue Anatomic Site •Pathologic Status	No	Open/Public
CGCC expression	•Gene Expression (raw and normalized)	No	Open/Public
CGCC methylation	•DNA methylation	No	Open/Public
CGCC raw SNP	•Raw genotype calls	Yes	DAC approval required
CGCC raw CGH	•Raw signal of hybridizing oligos	No	Open/Public
CGCC summary SNP and CGH	•Genotype frequencies •Computed Copy number •Loss of Heterozygosity	No	Open/Public
GSC mutation data	•Newly discovered somatic variants.	No	Open/Public
GSC linking table	•Information that links released data to the BCR complete set of clinical annotations	Yes	DAC approval required
GSC sequence traces	•Trace files with NCBI-required annotations. Traces from the same amplicon (forward-reverse reads) will be identified. Ability to aggregate all traces from a single sample across amplicons, however, will not be supported in the open/public data set.	No	Open/Public

Example: Overcoming Privacy Barriers

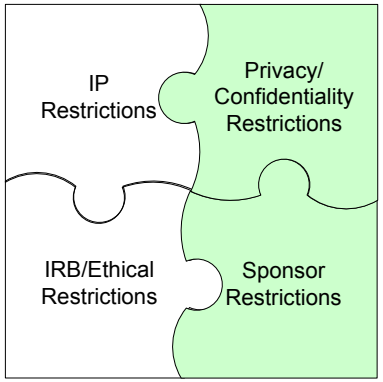
1. **Level of Identification**
   - Does the dataset to be shared include identifiable information?
   - If so, can identifiers be removed to create a "limited data set" or a "deidentified data set" without compromising the integrity of the research (*note: a deidentified data set may include links or codes to facilitate reidentification*)?
   - If identifiers cannot be removed, does disclosure meet another HIPAA exception
     - Review of decedents' information
     - Review preparatory to research (no data off-site -> inapplicable)
     - Waiver of authorization (granted by an IRB or privacy board)
2. **Protective Agreements: even if HIPAA (or applicable state law or institutional policy) restricts disclosure, restrictions generally can be addressed through use of appropriate agreements**
   - Deidentified data set: none necessary
   - Limited data set: data use agreement
   - Identifiable data: restrictive confidentiality agreements (not necessarily required from a federal regulatory perspective with subject authorization or if an authorization exception applies but practically important to assure subject protections and as industry "best practice")

Note: State laws and institutional policies can significantly impact this analysis. Many states impose special protections on genetic information, cancer information, behavioral health records, etc. Knowledge of these laws is essential to accurately identify privacy barriers and evaluate how best to overcome them.

TCGA: Privacy Restrictions (Federal)



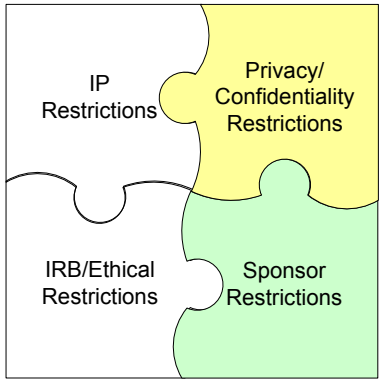
BCR Minimal Data Set



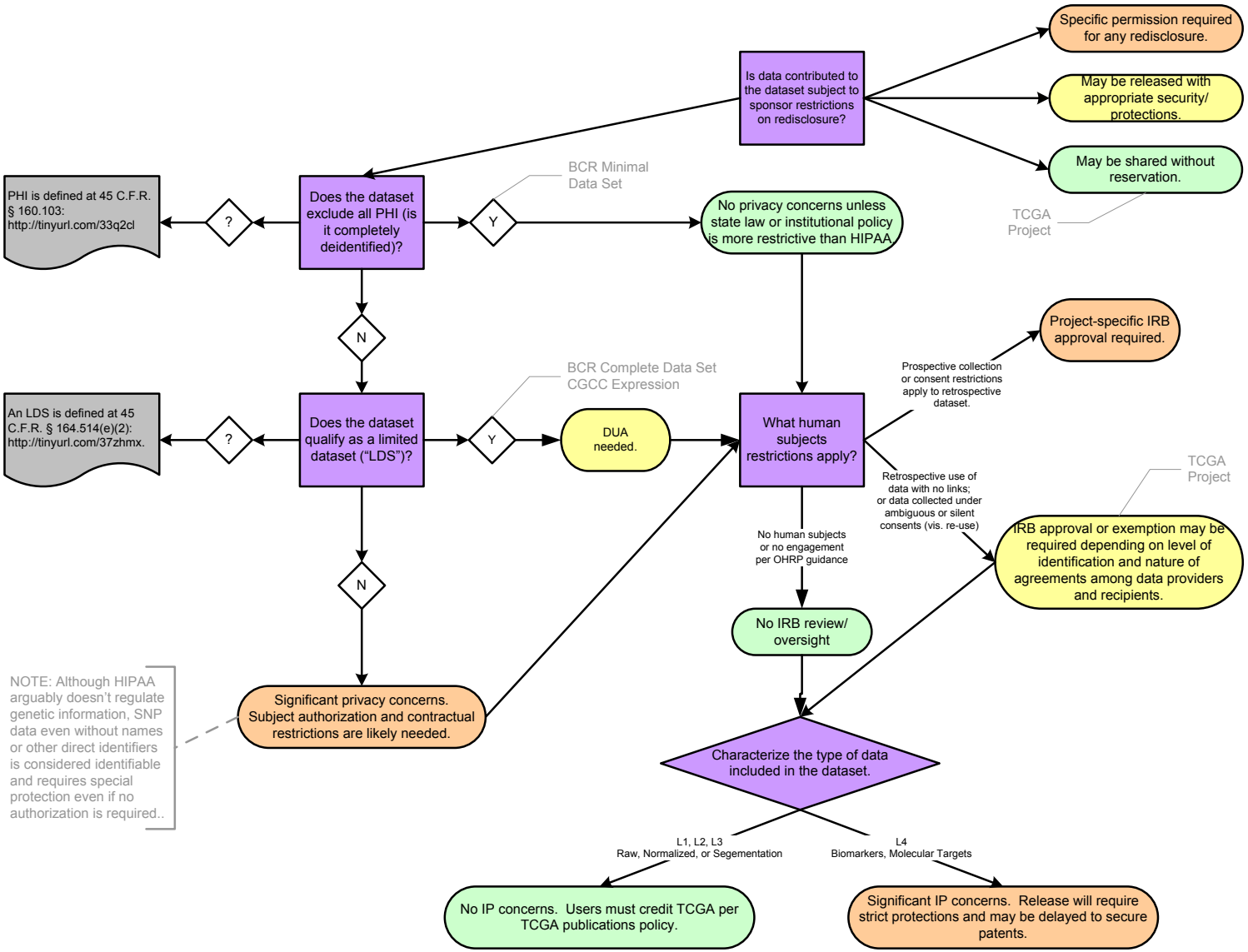
SNP Data*

*SNP data is not PHI but is considered Identifiable.

BCR Complete Set



Framework Applied



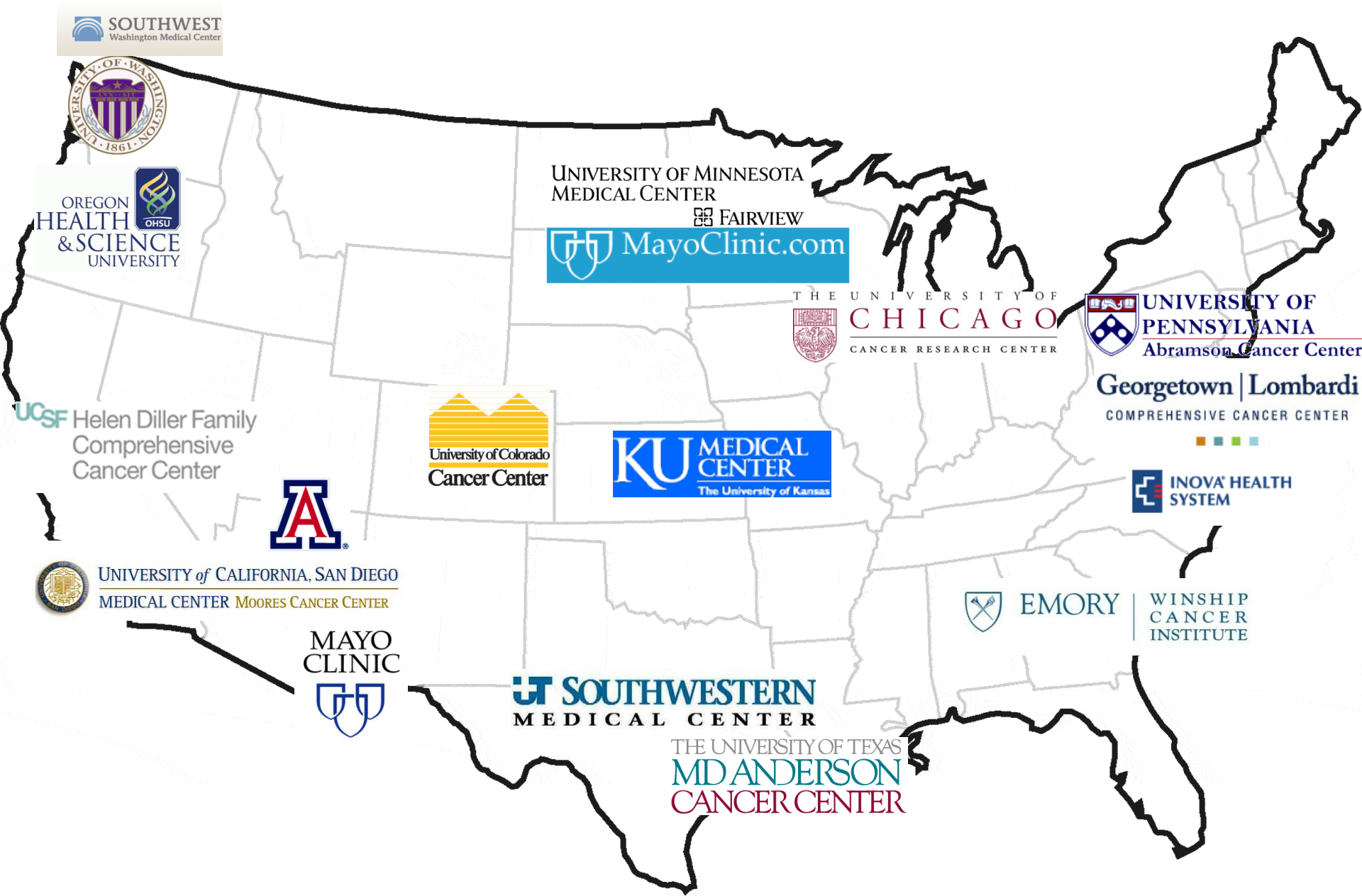
Case Study

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

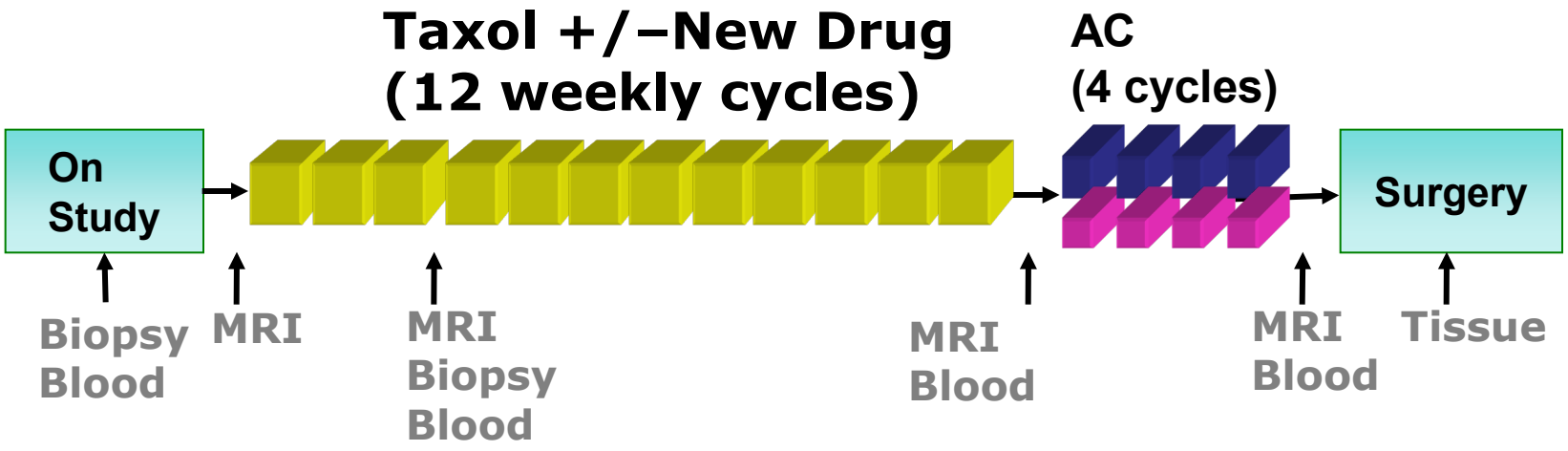
A national study to leverage biomarkers
in predicting response to combinatorial therapy for
women with Stage 3 breast cancer.

(PI Laura Esserman, UCSF)

Projected I-SPY 2 study sites



I-SPY Adaptive Trial Outline

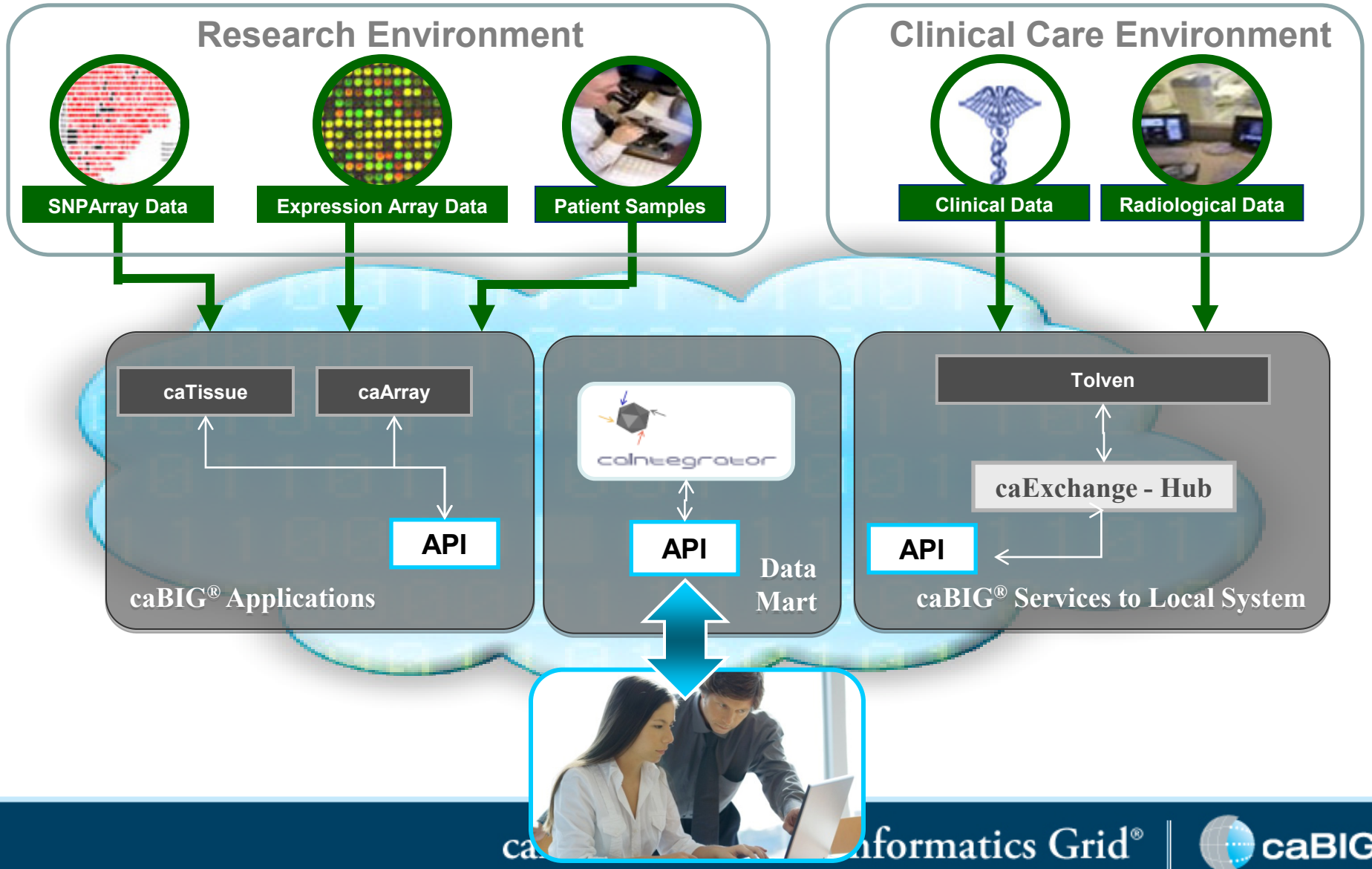


Accrual: Anticipate 800 patients over 3–4 years

Enroll ~20 patients per month

Participating Sites: 15–20 across US and Canada

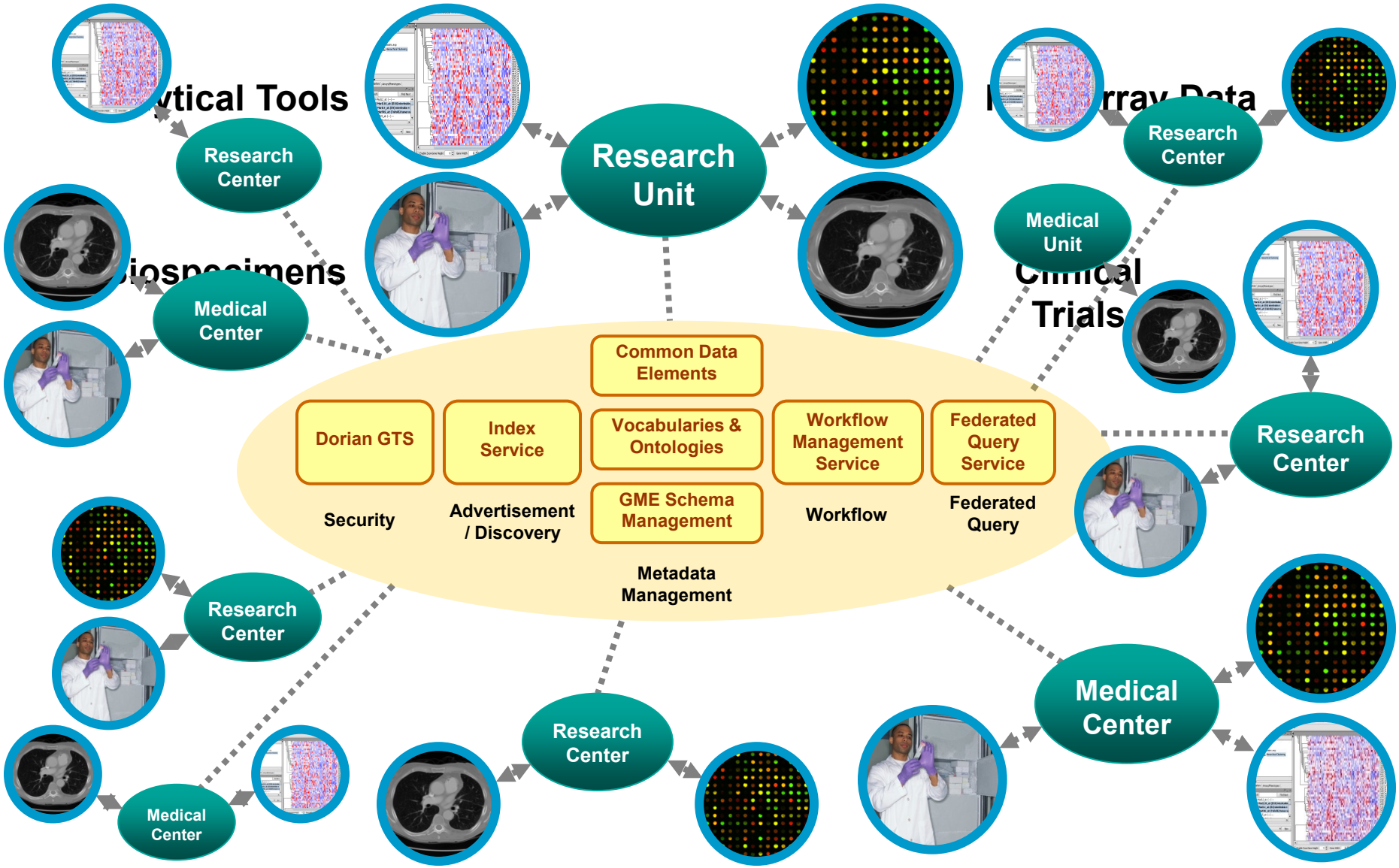
I-SPY TRIAL IT Infrastructure

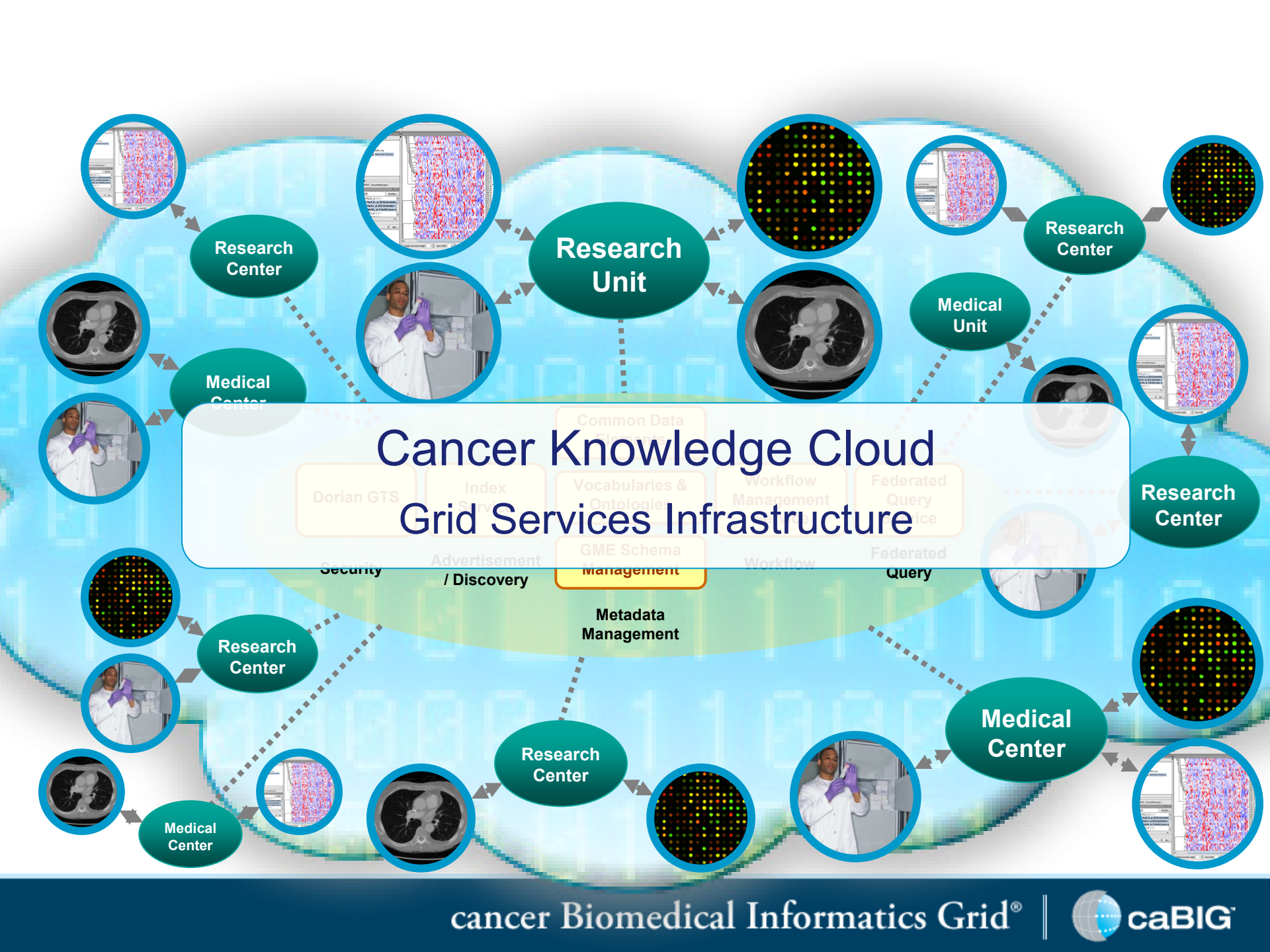


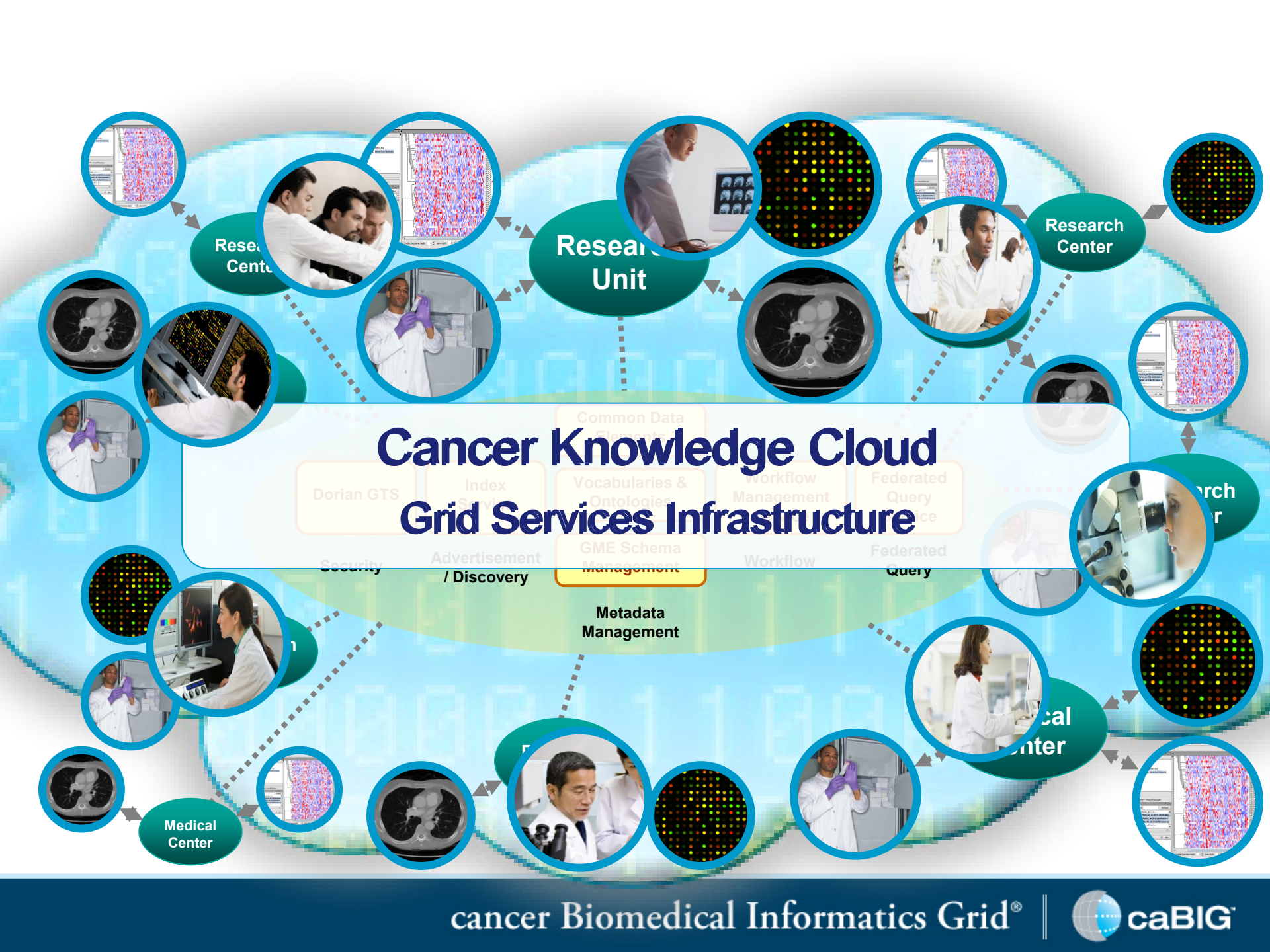
Future Direction of caBIG®

Future Directions

- **The external environment is advancing rapidly**
  - $44B through HITECH Act to digitize medicine
  - Rapid Learning Healthcare System
  - Globalization of research
  - Cloud computing
- **caBIG® 2.0 addressing these changes**
  - Linking research and care
  - Enabling genomically-subgrouped clinical trials
  - Enabling more diverse ecosystems of players


vtical Tools
Research Center
Research Unit
rray Data
Research Center
Medical Unit
iospecimens
Medical Center
Clinical Trials
Common Data Elements
Dorian GTS
Index Service
Vocabularies & Ontologies
Workflow Management Service
Federated Query Service
Research Center
Security
Advertisement / Discovery
GME Schema Management
Workflow
Federated Query
Metadata Management
Research Center
Research Center
Medical Center
Medical Center


Research Center
Research Unit
Research Center
Medical Unit
Medical Center
Research Center
Cancer Knowledge Cloud
Grid Services Infrastructure
Dorian GTS
Index Service
Vocabularies & Ontologies
Workflow Management
Federated Query Service
Security
Advertisement / Discovery
GME Schema Management
Workflow
Federated Query
Metadata Management
Research Center
Research Center
Medical Center
Medical Center



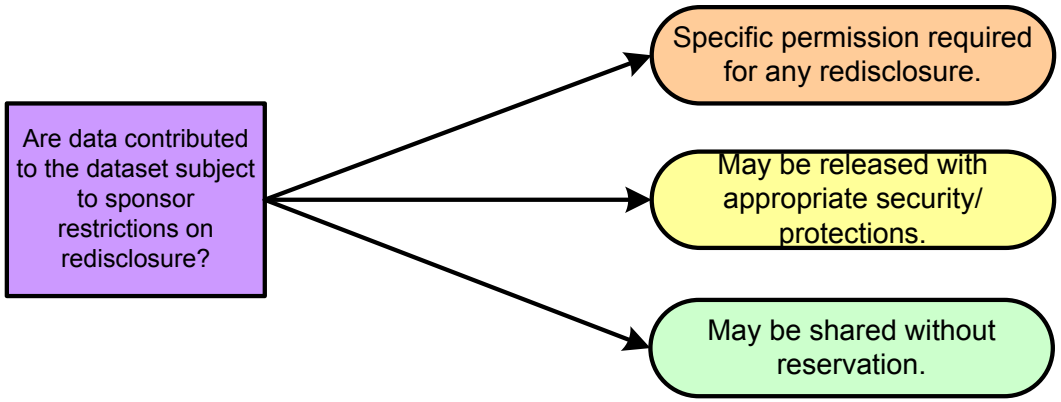
Cancer Knowledge Cloud

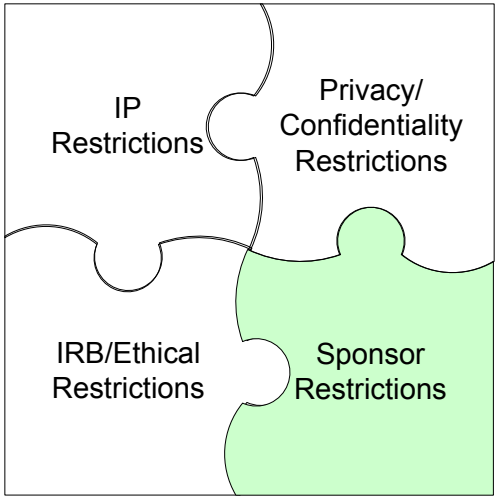
Grid Services Infrastructure

Questions?

Appendix

TCGA: Sponsor Restrictions

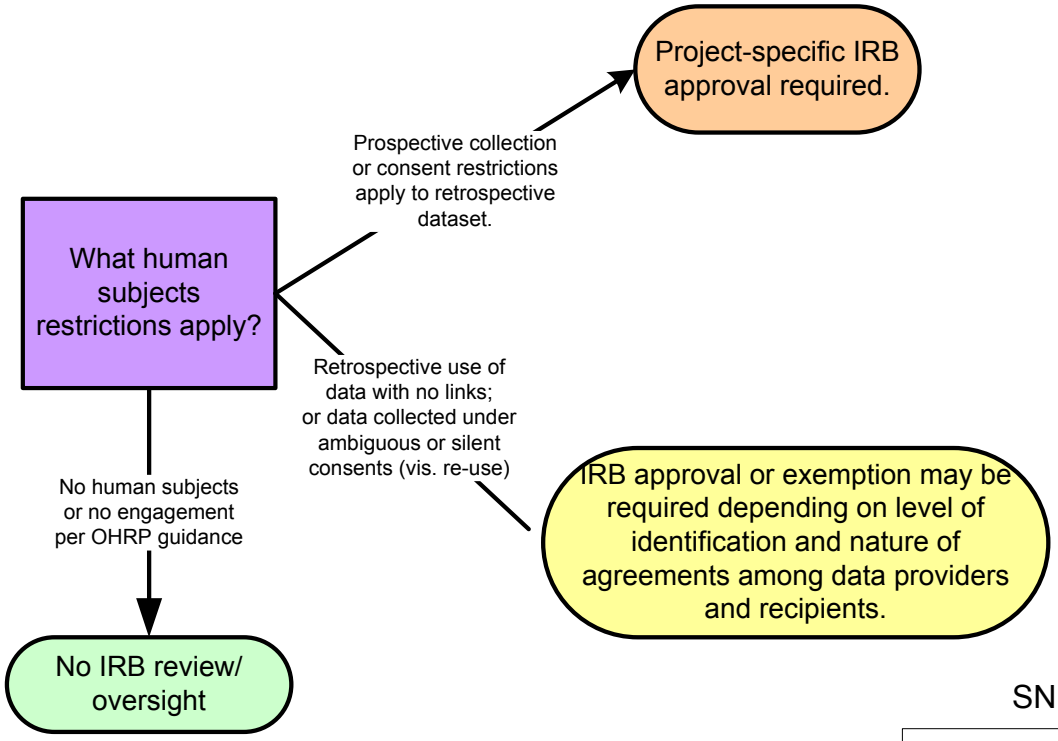




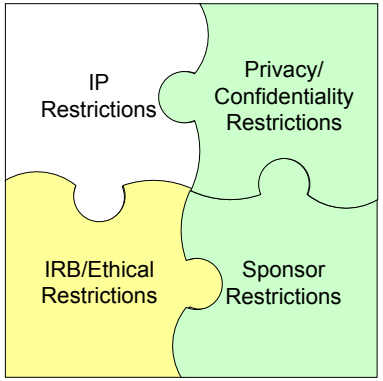
- **TCGA data/specimens were not collected under agreements with government, industry, or foundation sponsors that would prohibit use or disclosure via the caGRID.**
- **No special permission is required from sponsors to share TCGA data/specimens.**
- **No special agreements are required to share the data/specimens.**

Questions: When do sponsors impose moderate (e.g., publication delays) or substantial (e.g., classified research) restrictions on data sharing? What do these look like? Can data be shared at all under these circumstances? If so, how?

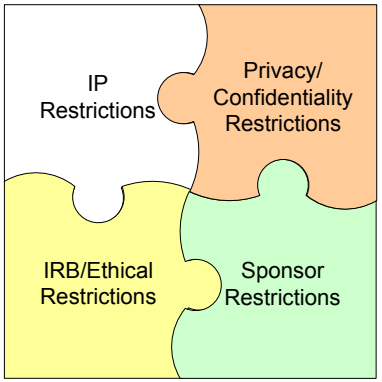
TCGA: IRB/Ethical Restrictions (Federal)



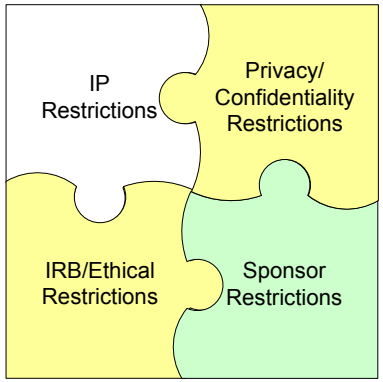
BCR Minimal Data Set



SNP Data



BCR Complete Set



- Participant IRBs have granted “exemptions” from IRB oversight because the research involves use of existing data that cannot be linked back to individual subjects.

Overcoming IRB Barriers

IRB Oversight

1. **Distinguish human subjects research from unregulated research**
   - Human subjects are alive
   - OHRP has issued guidance on "coded information and specimens" (http://www.hhs.gov/ohrp/humansubjects/guidance/cdebiol.htm) and draft guidance on institutional "engagement" (http://www.hhs.gov/ohrp/requests/engage.html) that together define many activities currently regulated by IRBs as non-human subjects research
   - In most institutions, non-human subjects research is not subject to IRB oversight, though who makes that decision with respect to any given project varies

2. **Determine whether the research is "exempt" or whether a proposed project or inquiry is covered under a "master" or "umbrella" protocol**
   - Studies involving previously collected data that cannot be directly or indirectly linked to living individuals are eligible for exemption, which generally must be granted by the IRB
   - IRBs may sometimes approve "master" or "umbrella" protocols that cover a broad range of individual projects or analyses

3. **For non-exempt human subjects research, consider alternatives to multi-institutional approval**
   - CIRB
   - Commercial IRB
   - Defer or accept review under an IRB Authorization Agreement (http://www.hhs.gov/ohrp/humansubjects/assurance/iprotsup.rtf)

Overcoming IRB Barriers

Informed Consent

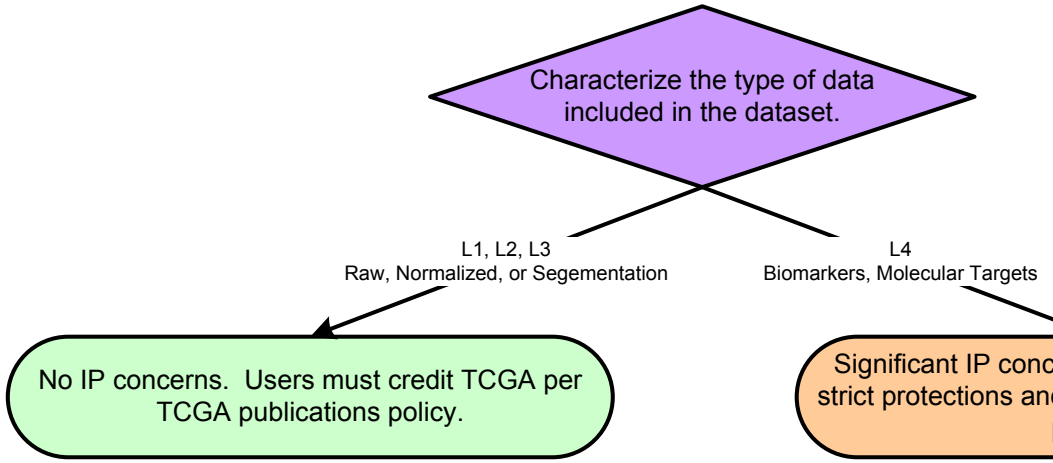
1. **Explicit permission to share data (and specimens) with researchers via caGrid (or more broadly through a data/specimen registry) can help eliminate IRB or broader ethical barriers**

2. **Absent explicit permission, IRBs may permit retrospective research on data or specimens previously collected under clinical or research consent documents that were silent or ambiguous about future use for unspecified analyses conducted by the original research team or others**
   - Many older consents include explicit language that restricts use of data or specimens to the current project
   - Response varies: IRBs may permit reuse under a waiver if deemed consistent with original intent of the consent, may permit re-contact with subjects to solicit explicit permission, or may bar reuse and re-contact

3. **Explicit permission (or IRB-approved waiver) is required for prospective collection of data or specimens**
   - Waiver may be difficult to secure because it requires a showing of "impracticability"

We will have an expanded discussion on this topic later today . . .

TCGA: IP Restrictions



L1 = Raw Data

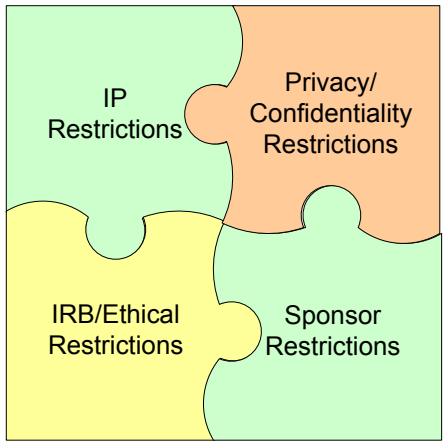
L2 = Normalized Data

L3 = Segmentation Data

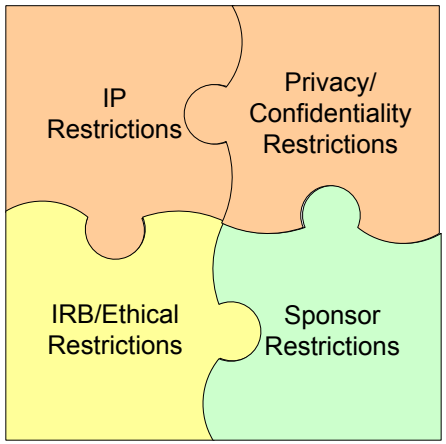
L4 = Biomarkers, Molecular Targets

CGCC Raw SNP

L1-L3



L4



Tomorrow: What "IP" concerns drive data sharing restrictions imposed by research institutions or individual researchers? How can these be addressed to facilitate data sharing?

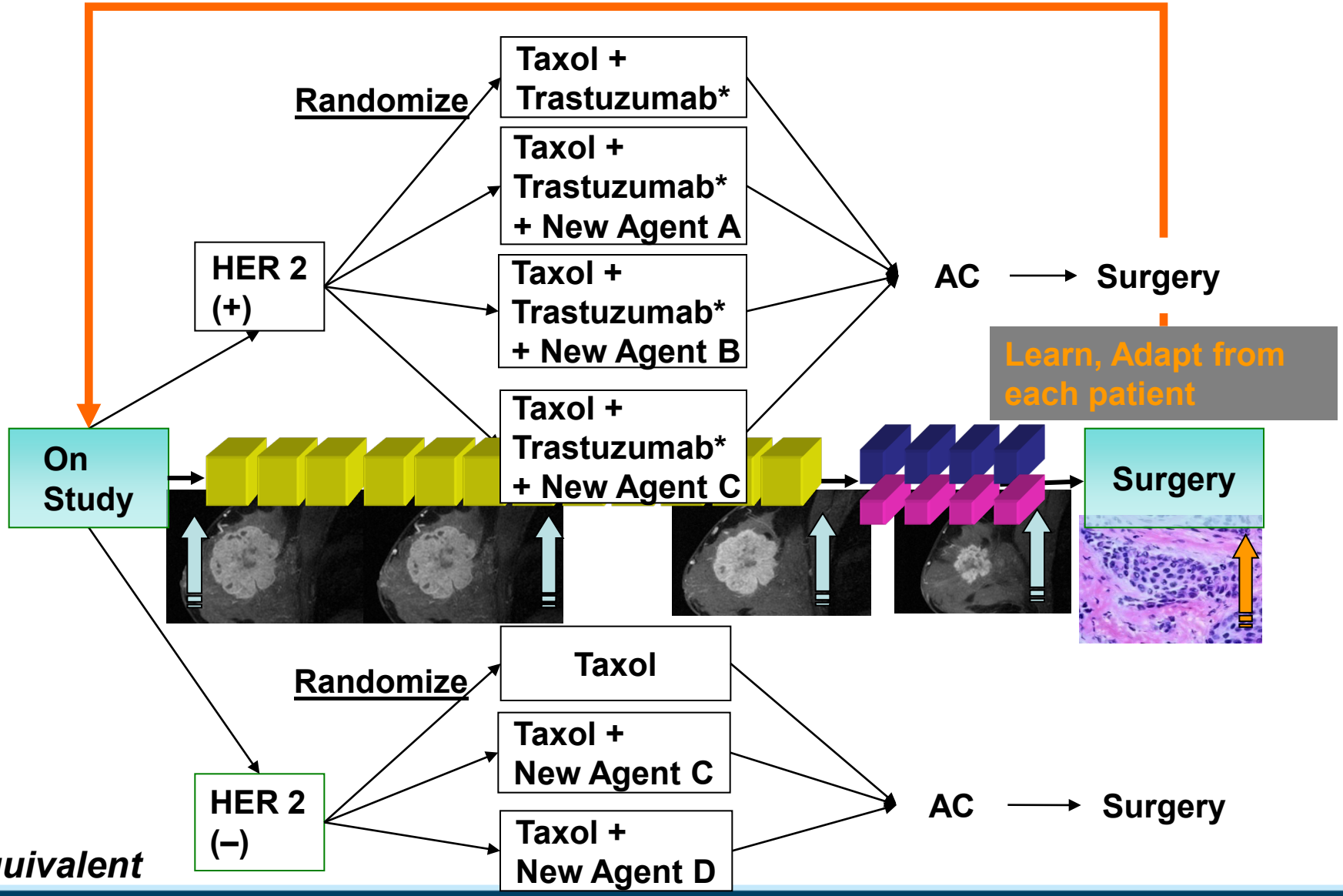
The “big picture” . . .

- **The need to protect key rights or interests presents challenges to data sharing in caBIG environment; DSIC WS thinks about these issues in separate “buckets”:**
  - ***Regulatory SIG – privacy, autonomy . . .***
  - ***Proprietary SIG – intellectual property, publication . . .***

- **Security policies and procedures for sharing data via caGrid technology stack are the mechanism to protect these rights or interests; NCI-caGrid Security Working Group (SWG) develops recommendations for security policies and procedures**
  - ***DSIC Regulatory SIG supports SWG on security policy matters***
  - ***Architecture WS supports SWG on technical security matters***

I-SPY Adaptive Trial:

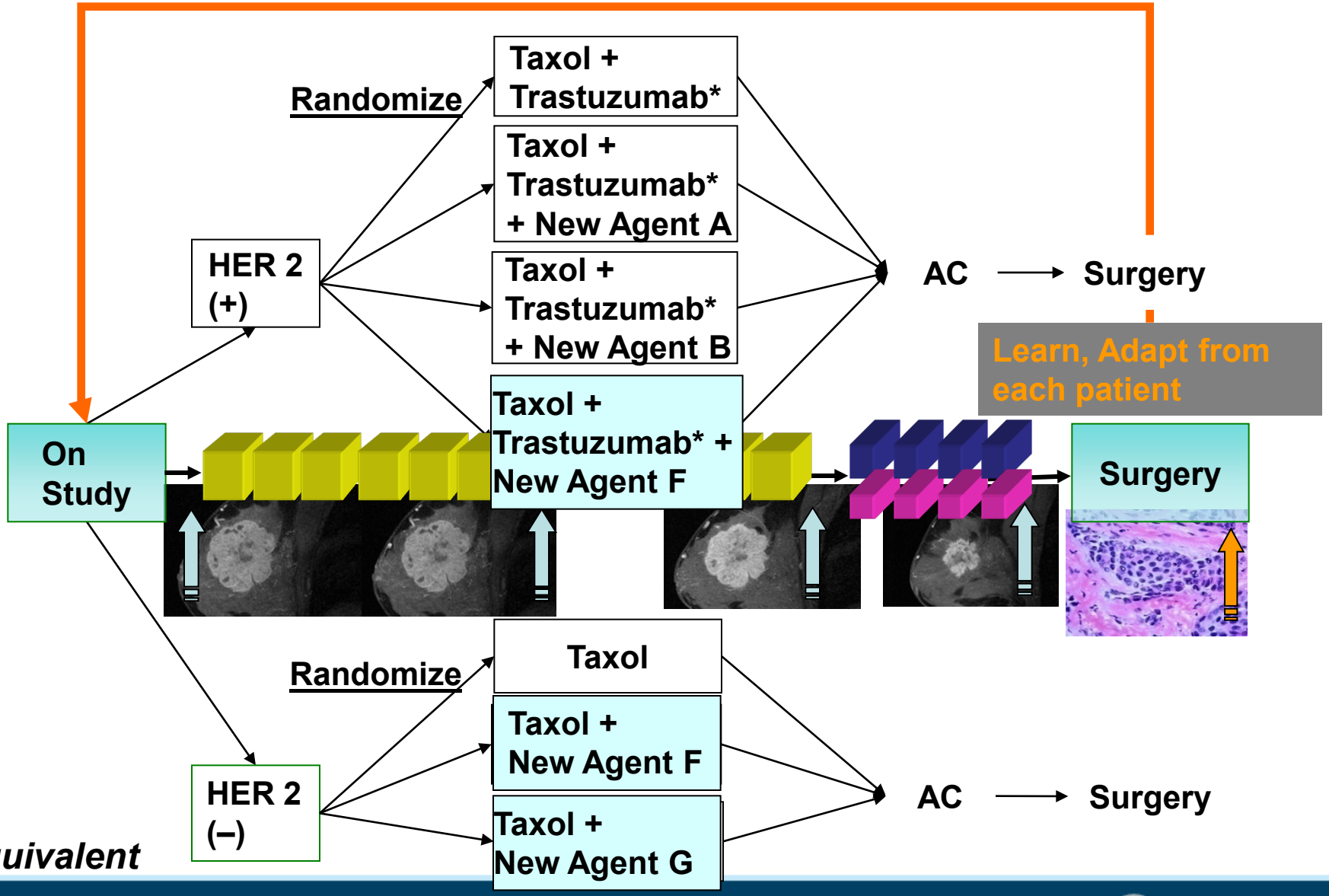
Introduce several new agents for a given profile



**Or Equivalent*

I-SPY Adaptive Trial:

Introduce several new agents for a given profile



**Or Equivalent*

Cancer Center Cooperative Development Meetings

- **Goal: collect qualitative data to understand Centers' bioinformatics goals, their progress and challenges in meeting those goals, and how caBIG® has contributed to success/accomplishments**
  - Identify how Centers have made use of caBIG® tools and infrastructure
  - Identify enterprise-level requirements for future success in informatics, including workflow and system interconnectivity needs
  - Identify ways in which NCI may facilitate use of caBIG® tools, infrastructure and policy
- **Three phases:**
  - Selection: which Centers to interview
  - Content: what questions to be asked
  - Collection: visits, interviews, data and analysis