NCI caBIG biomedical informatics

- **Goal**: A virtual web of interconnected data, individuals, and organizations redefines how research is conducted, care is provided, and patients/participants interact with the biomedical research enterprise

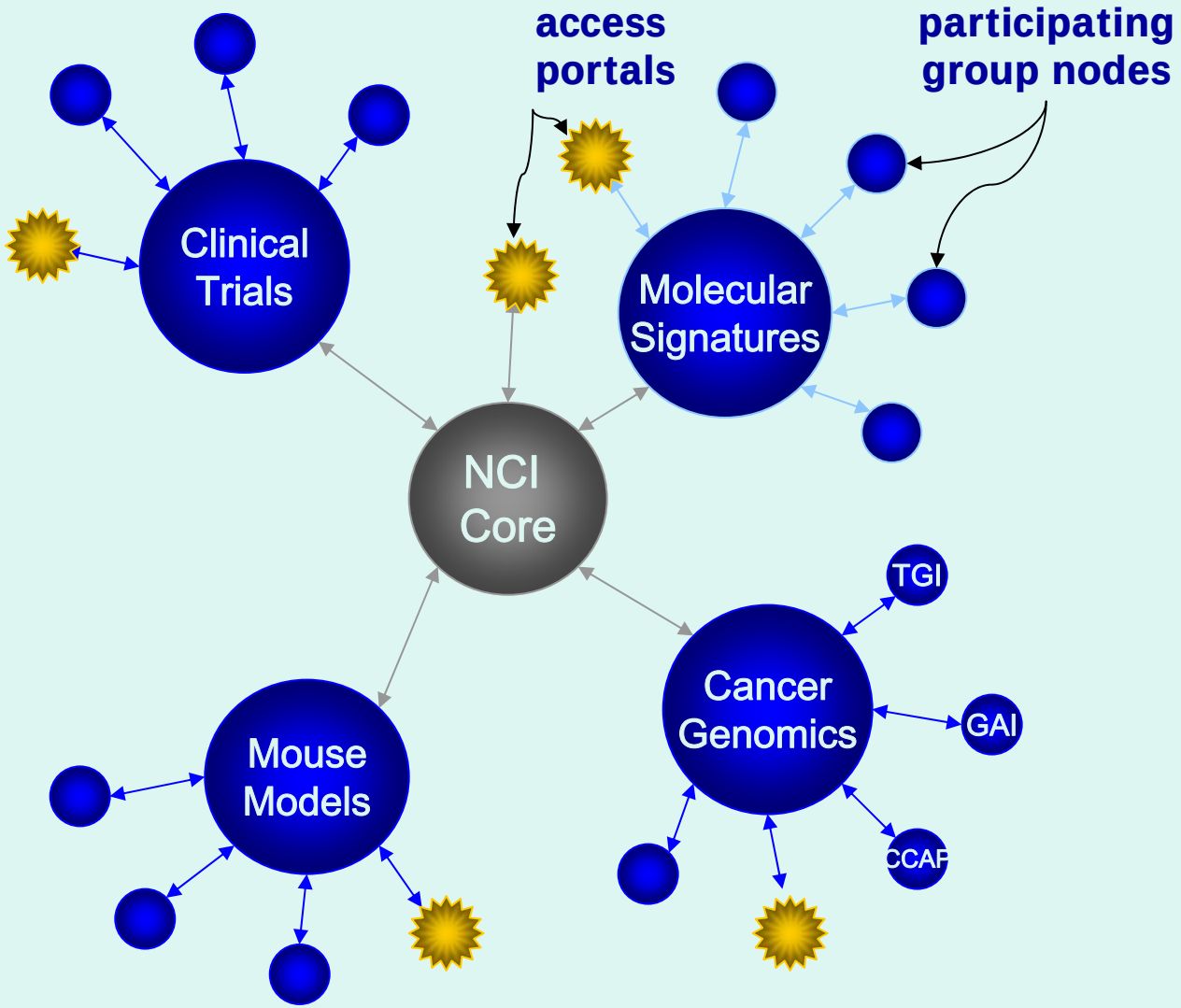
NATIONAL CANCER INSTITUTE

- Common, widely distributed infrastructure permits cancer research community to focus on innovation
- Shared vocabulary, data elements, data models facilitate information exchange
- Collection of interoperable applications developed to common standard
- Cancer data is available for mining and integration

caBIG consortium

- NCI Cancer Centers
- SPOREs
- Intramural Program
- Specific Initiatives
  - COOP Groups
  - CIP
  - MMHCC
- Other biomedical research groups and consortia

building common architecture, common tools, and common standards

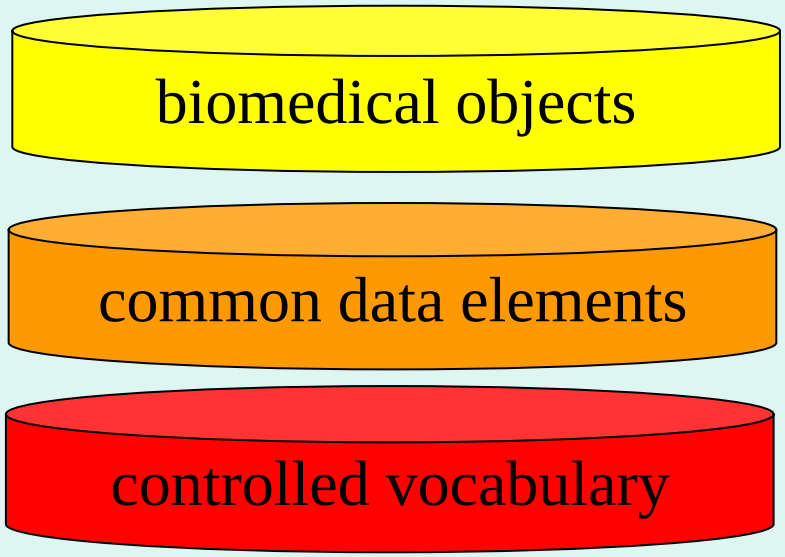


caBIG principles

- Open source
- Open access
- Open development
- Federated

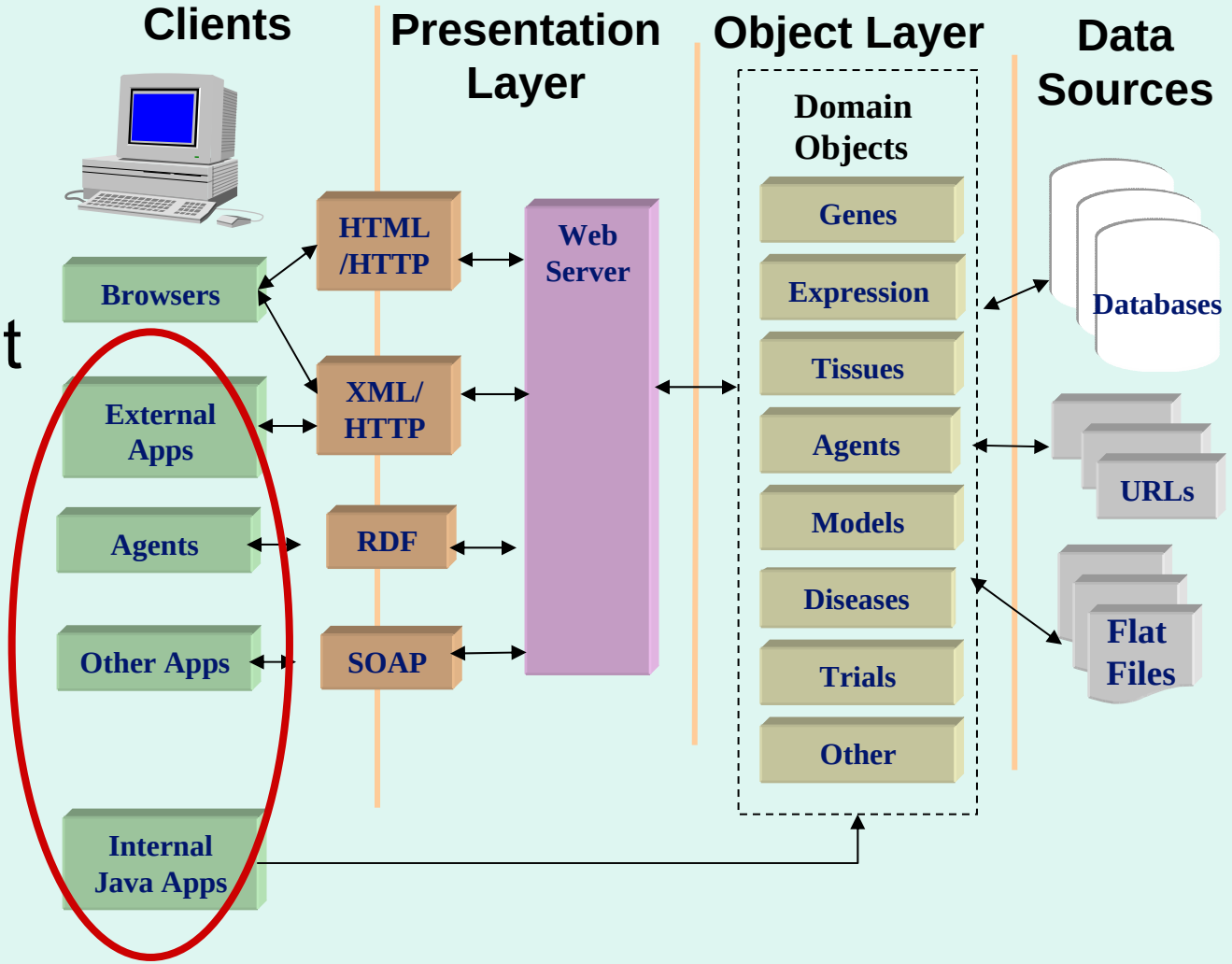
caCORE – common ontologic representation environment

- Information integration
- Cross-discipline reasoning



…a sharing architecture

- open source
- open development
- open access
- federated



NCICB applications:

- clincial trials support - C3D
- molecular pathology - caArray
- cancer images - caImage
- pre-clinical models - caModelsDb
- laboratory support - caLIMS

caBIG action plan

- Establish pilot network of ~10 Cancer Center
  - Groups agreeing to caBIG principles
  - Mixture of capabilities
  - Mixture of contributions
- Expanding collection of participants
- Establish consortium development process
  - Collecting and sharing expertise
  - Identifying and prioritizing community needs
  - Expanding development efforts

CIP effort: Dancing with the elephant

- **I2 "<u>Integration & Implementation</u>" Initiative**

Today's Imaging Cancer Research & Practice = Undeveloped Potential

- **Detection**
  - Tedious
  - Inefficient
  - Observer-dependent
  - Not reproducible

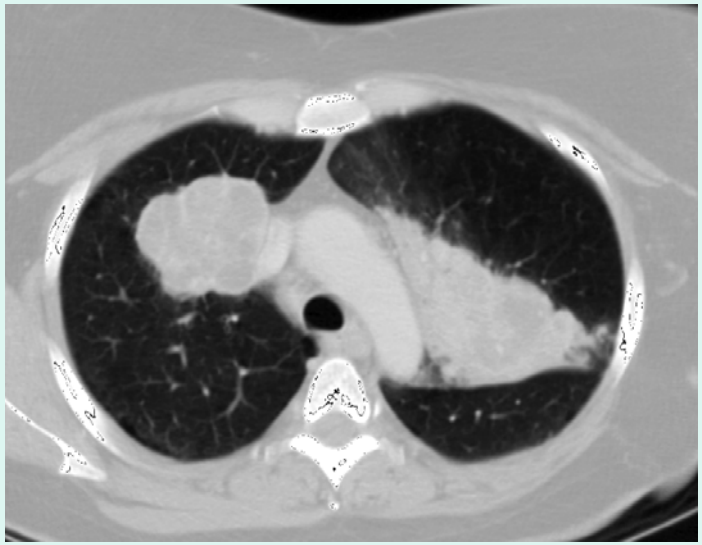
  → Automated / semi-automated; Efficient; Less observer-dependent; Reproducible
- **Diagnosis**
  - Inaccurate
  - Invasive confirmation required

  → Accurate classification; Noninvasive
- **Response to therapy**
  - RECIST criteria inadequate
  - Delayed decisions on individual care

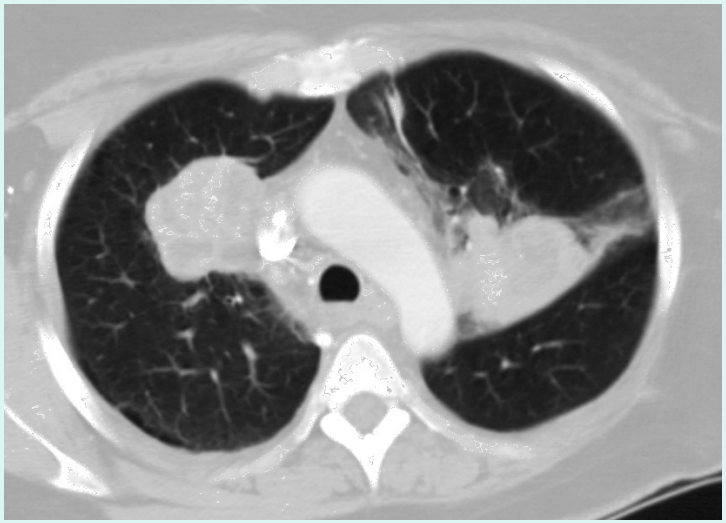
  → Improved objective measures; Early individual care decisions
- **Development of new therapeutics**
  - Drug studies lack imaging surrogates
  - Trials costly; too many patients
  - Long path to market & patient benefit

  → Early go/no-go decisions; Trials less costly, fewer patients; Shorter path to market

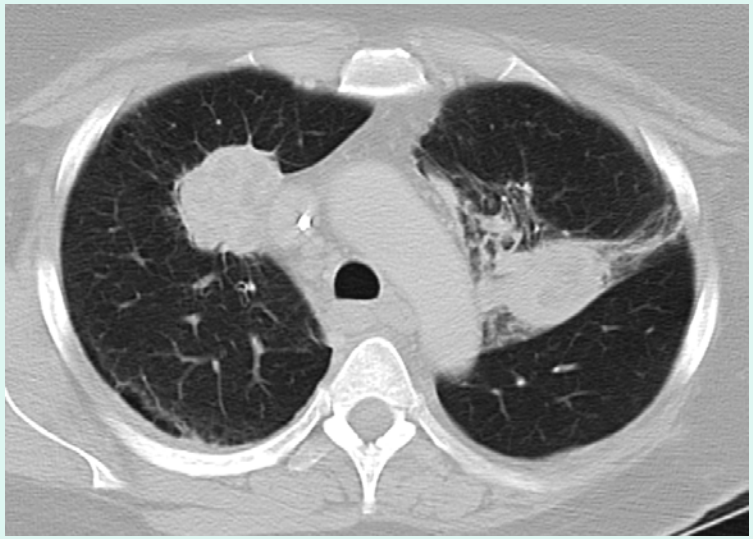
Baseline



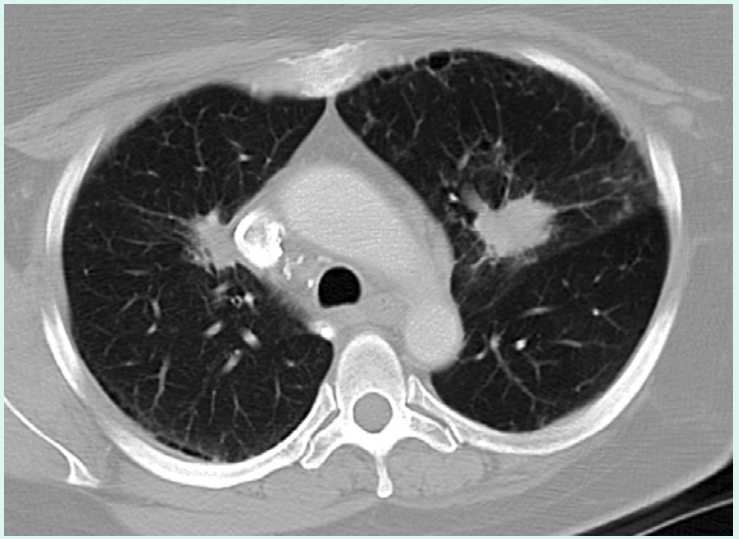
20 weeks (PR at - 39%)



24 weeks (PR confirmed - 52%)



52 weeks (- 74%)



metastatic renal cell carcinoma

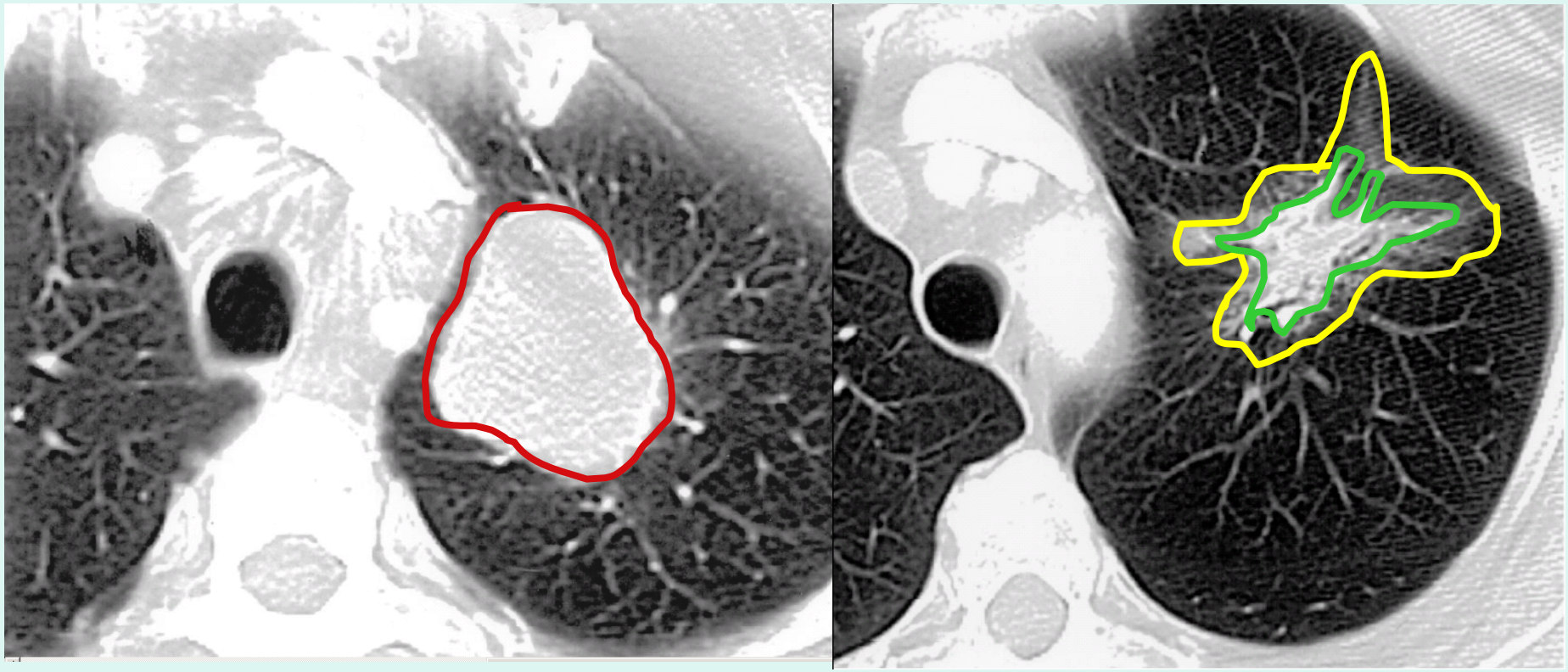
Cost to conduct a study*

Comparison Between SURVIVAL & TTP Studies

Therapy	Sample Size		Trial Cost*	
	Superiority	Non-inferiority	Superiority	Non-inferiority
SURVIVAL STUDY CPT-11/5-FU/LV ⇒ Oxali/5-FU/LV	**2200**	**5700**	**$88M**	**$228M**
TTP STUDY CPT-11/5-FU/LV ⇒ Oxali/5-FU/LV	**400**	**750**	**$16M**	**$30M**

*This example from TP Therapeutics on colon CA, assumes a costs of ≈ $40,000/patient

Barrier to consistent data



Critical Path

- Make image data trustworthy. How?
- Validated analytic software tools for:
  - Lesion detection, classification
  - Accelerated diagnostic imaging decision throughput
  - Quantitative imaging assessment of drug response
- Missing ingredient: Image Database Resources

Statement of the Problem: Current Business Model

- Industry must individually acquire image databases for software applications such as screening, diagnosis, or image-guided intervention.

  Model relies on partnerships with academic sites to access images from clinical trials; data accrual and content from these trials are often not suitable for software validation.

CIP Premise

Image processing software and image archives for software validation are urgently needed for:

- Detection, Classification,
- Quantitative Monitoring, Rx response
- Accelerating and standardizing FDA drug approval

Communities to be served:

- Academic medical and computer science researchers
- Device and Drug Industry <- a significant, under-served constituency
- FDA CDRH

CIP Principles

- Leverage relationship with imaging professional organizations to address critical cancer needs
- Foster Inter-institutional, inter-agency alliances
  - NLM's ITK; FDA's IAG's; Navy's CTC; etc..
- Develop consensus on structured, standardized exchange and use of information

CIP Goal:

Informatics to optimize value of cancer imaging data.

Major Objectives:

Establish publicly available image archives, linked to outcome and other clinical data

Stimulate development and dissemination of open-source image processing (e.g. CAD) software

Partner with industry (FNIH) to support image archives
Public Private Partnerships with the device and drug industry

NCI's Capabilities

- **NCI sponsors a range of clinical trials that involve imaging methods for cancer screening, diagnosis and therapy response (ACRIN, Cancer Centers, SPOREs)**

- **NCI-funded investigators are developing databases for software validation that can be integrated into this initiative.**

- **NCI has established a bioinformatics support group that includes an image and data archive for new clinical trials, with secure web-accessible queries.**

Proposed Business Model

- **Develop industry partnerships with NCI, academia and the FDA; coordination by the FNIH.**
- **Pool resources from industrial partners that have a common interest in software validation for cancer applications.**
- **Form steering committee(s) with national or international representation of all parties and formulate plans for database collections of interest to both academia and industry, including timetables for access.**

Proposed Business Model

- **Develop a broad-based consensus for the underlying science involved in database development that may lead to more objective and standardized methods for software validation.**

- **Provide secure web-based access to the databases.**

- **NCI will coordinate the effort and integrate it with NCI's plans for an image archive and informatics infrastructure.**

Business Model: Leveraging of Resources

- **Clinical Data Collection:** Costs covered by on going clinical investigations. Support is required for archiving of data sets and related image annotation required for software performance assessment.
- **CaBIG**: Web accessible methods to query this public resource are being developed as an integral part of the caBIG " *Imaging Workspace*"
- **Public Private Partnerships:** Engage cancer center, academic, and the device and drug industry communities to develop and support public databases. Includes FDA and NIST scientists with a goal of using this database as a resource to accelerate regulatory approval, standardized assessment of informatics tools by industry, and reimbursement by CMS.
- **Developers:** Engage the broader scientific community to develop more advanced software tools without concern about data collection/annotation.
- **Physician End Users:** Encourage this community RSNA (IHE) to require more standardized methods for software evaluation so that informatics tools will be widely accepted by the radiology community.

REFERENCE IMAGE DATABASE to EVALUATE RESPONSE
to Drug Therapy in Lung Cancer

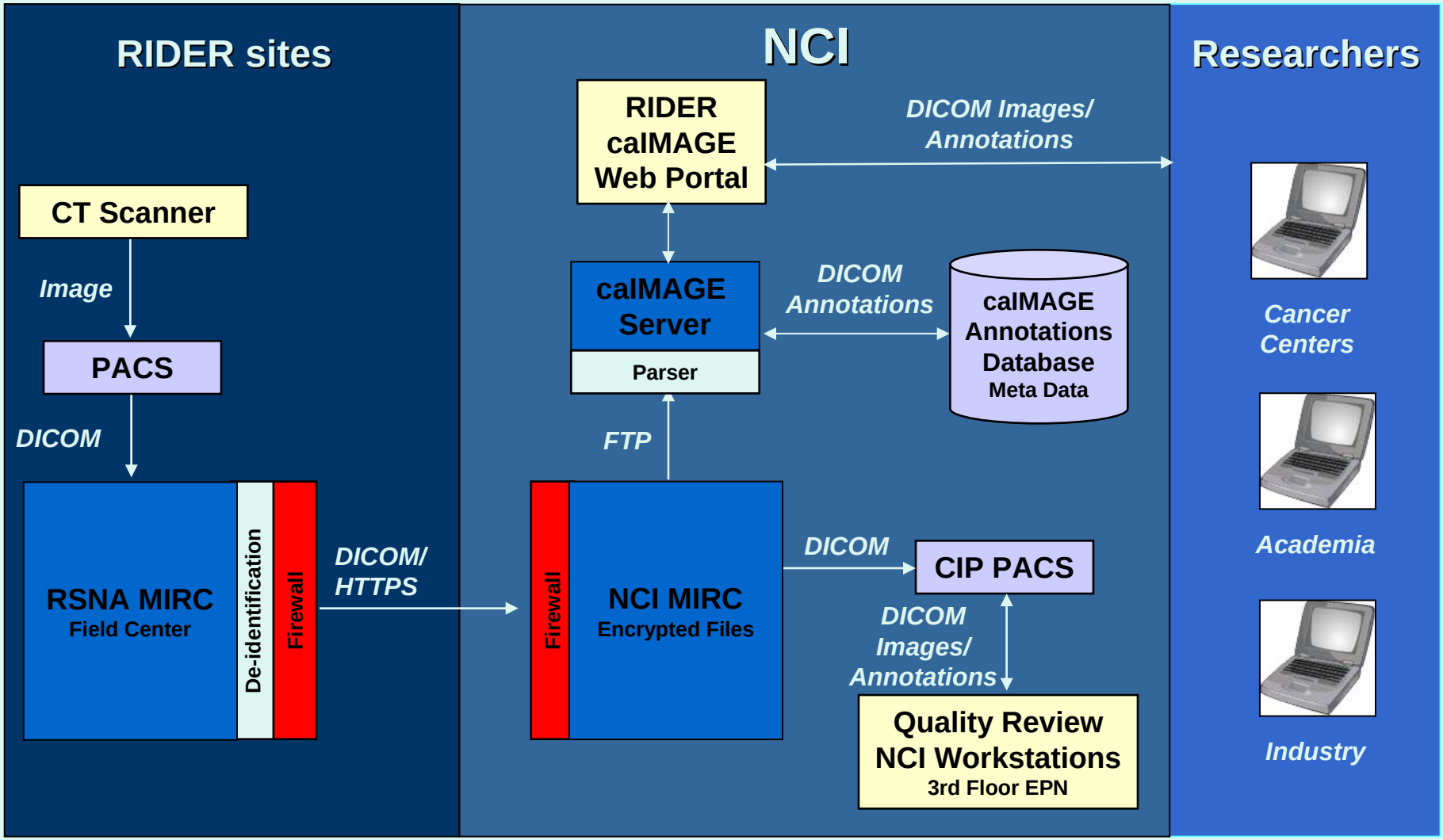
RIDER Aims:

- *Pilot* database: 200 advanced lung CA patients; serial CT exams.
- A step toward NCI imaging informatics infrastructure
- Expert consensus on database design
- Enable industry & academia to develop, test, compare semi-automated, automated software tools for change analysis
- Dovetail RIDER with larger image database initiatives
- Aid partnering between FNIH, NCI/NIH, FDA, industry, academia to support future database resources (public-private partnerships)

REFERENCE IMAGE DATABASE to EVALUATE RESPONSE

- Issues being addressed in RIDER
  - Harmonization of imaging protocols
  - Firewalls
  - Privacy (HIPAA, deidentification / anonymization)
  - Access
  - CaBIG Compatibility (linkage to CA research communities and data)
- Issues still to be addressed in RIDER
  - Database design consensus
  - Metadata to include
  - Quality review process
  - Curation
  - Pilot evaluation of software tools by researchers on RIDER data subsets

RIDER Infrastructure



REFERENCE IMAGE DATABASE to EVALUATE RESPONSE
to Drug Therapy in Lung Cancer

- **Investigators:**

  1) LIDC principal investigators (U01)

• U of Iowa	McLennan
• U of Chicago	Armato
• U of Michigan	Meyer
• Cornell U	Yankelevitz
• UCLA	McNitt-Gray

  2) Cancer Center investigators (P30)

• MSKCC	Schwartz
• MDACC	Munden

- **Steering Committee:** 2 members per site, L. Clarke, Chair, Barbara Croft PD, other NCI staff, FDA CDRH, NIST IT.

- **NCI** (CIP, NCICB); **RSNA**; **Contractors** (SAIC, TerpSys); **NIBIB**

CIP Near Term Goals

Potential for near term success

Develop validated* data collections:

- Lung nodules (FNIH Demonstration Project)
  - for Detection, Classification, rx. Response
- Liver mets - rx response
- Colon polyps - screening detection, classification
- Breast DMIST - detection, classification

*validated = image-marked up overlay + pathology +/- lab data

Timeline (3 months)

- **Enlist industry representatives and establish the steering committee.**
- **Engage NCI cooperative groups to enable data distribution responsive to this initiative.**
- **Initiate the first demonstration project.**
- **Expand NCI informatics and image archive infrastructure to meet the needs of this initiative.**

Timeline (1-3 years)

- **Complete a demonstration project:**
  - **Single modality database with > 1,000 subjects**
  - **Images linked to demographics, clinical data, and interpretations / reports**
- **Satisfy requirements of ACRIN, Cancer Centers, SPOREs etc.**
  - **Data access, security, confidentiality, investigator rights to enable project expansion.**
- **Distribute the database to industrial partners and assist in regulatory processes for new CAD product(s).**