caBIG™ cancer Biomedical Informatics Grid ™

An Initiative of the National Cancer Institute

In vivo imaging, caGrid and Cooperative Groups

Joel Saltz, MD, PhD

Chair Biomedical Informatics

Ohio State Comprehensive Cancer Center

College of Medcine

The Ohio State University

saltz@bmi.osu.edu

NIH Natcher Center October 18, 2007

Roadmap

- **Middleware**
- **Images, Cooperative Research Groups and the Grid**

Middleware: What Needs to be Done

- **Standard data representation**
- **Standard data meaning**
- **Tools for linking data and compute services**
- **Security that controls access to data and compute services to authorized groups**
- **Making sense out of (large) data**
- **Dynamic data driven application systems**

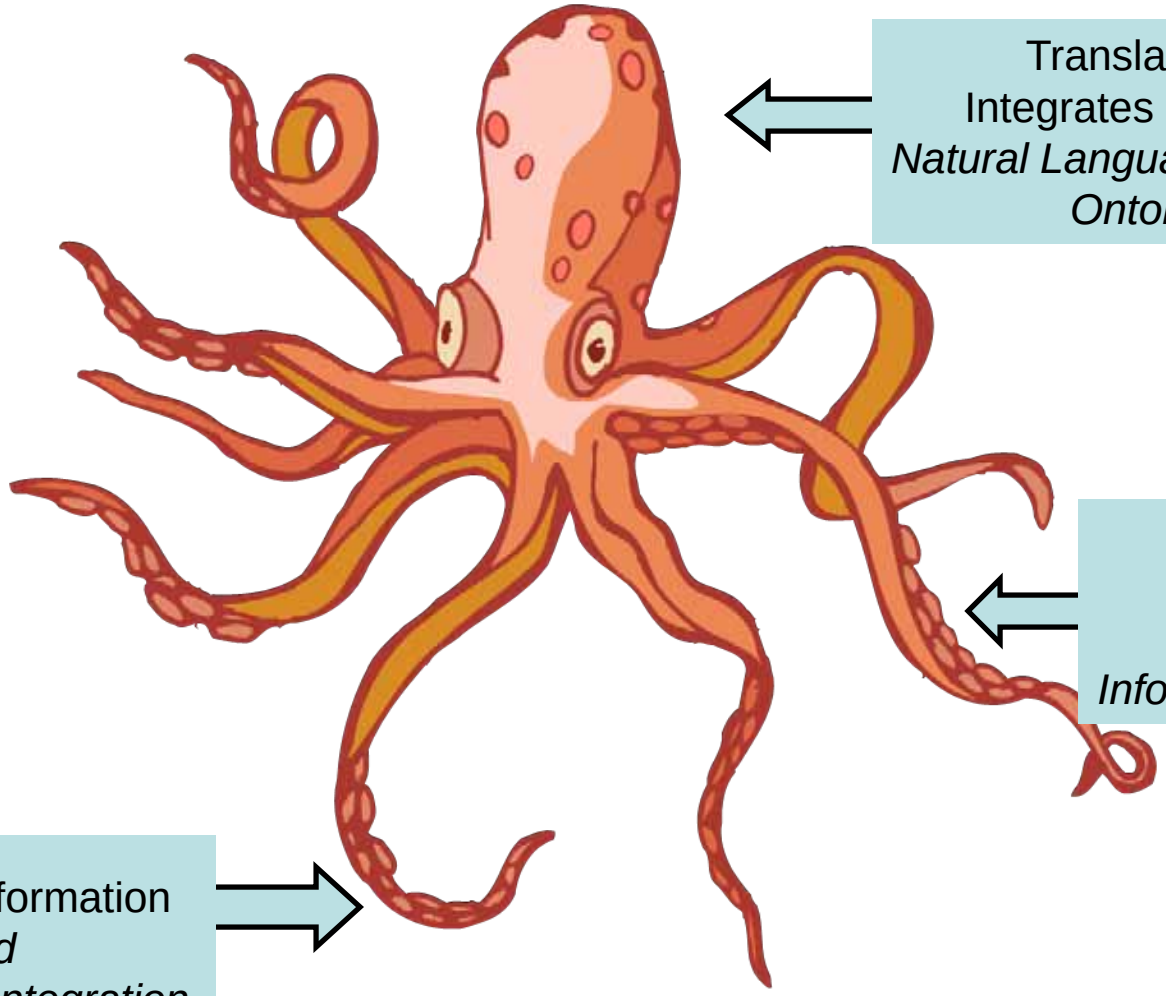
Transcontinental Railway:
The Golden Spike - Triumph of Standards

caBIG cancer Biomedical Informatics Grid™

OHIO STATE UNIVERSITY Medical Center

Translation: Same ideas, different words

Biomedical Informatics and Middleware



Translates and
Integrates Information
Natural Language Processing
Ontologies

Disseminates
Information
Grid
Information Integration

Brings in Information
Grid
Information Integration

Middleware: What Needs to be Done

- Standard data representation
- Standard data meaning
- Tools for linking data and compute services
- Security that controls access to data and compute services to authorized groups
- Making sense out of (large) data
- Dynamic data driven application systems

caGrid let's you…

- Discover data and analysis services
- Advertise your own services
- Invoke services
- Orchestrate Workflows
- Apply grid-wide security and access controls

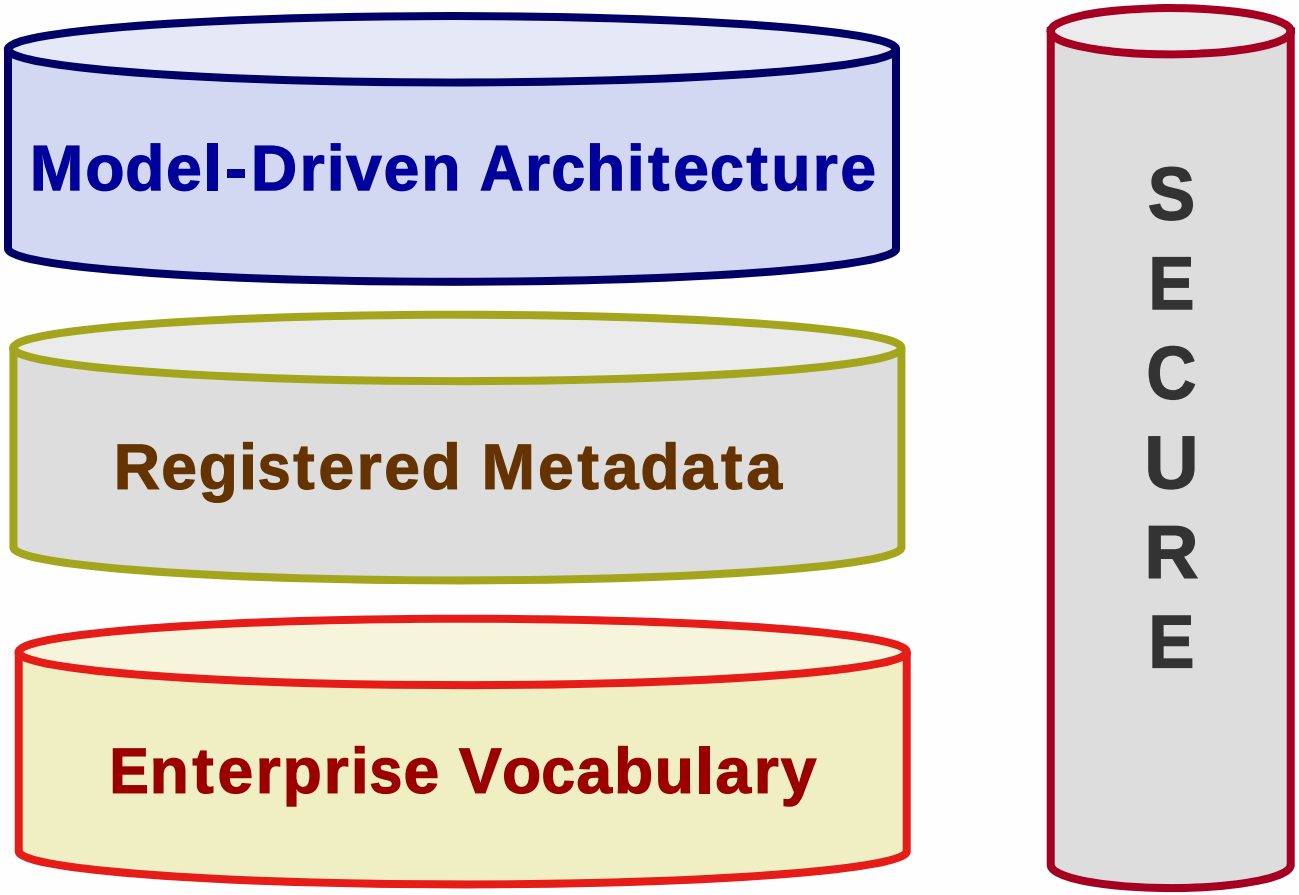
caGrid: Service-Oriented Data and Analysis Grid

- caCORE
  - EVS Terminology services
  - caDSR Metadata registry
  - Model-driven architecture Software Development Kit
- Mobius
  - Global Model Exchange (GME) XML infrastructure
- Globus
  - Standards-based toolkit for grid computing

- Globus-based data and analytical service grid
- Strongly typed data objects
- Registered data models
- Grid-wide object identifiers
- XML transport format
- Federated security and identity management
- Rapid service creation tools

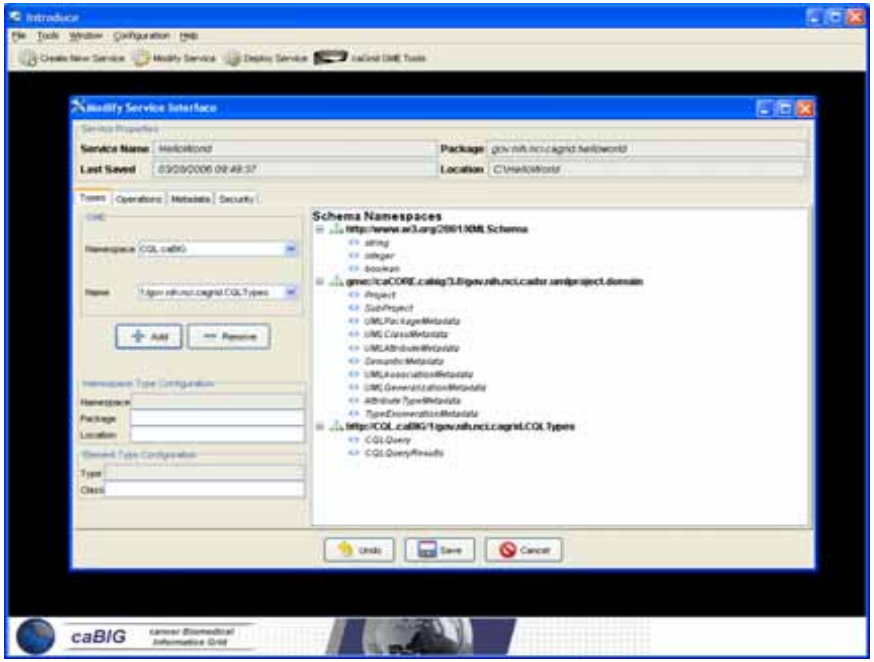
Medical Center

caCORE


Model-Driven Architecture
Registered Metadata
Enterprise Vocabulary
SECURE

caGrid: Core and IVI Middleware Extensions

- **caGrid Components**
  - Language (metadata, ontologies)
  - Security
  - Advertisement and Discovery
  - Workflow
  - Grid Service Graphical Development Toolkit (Introduce)
  - Efficient Bulk Data Transport (IVI middleware)
  - DICOM compatability (IVI middleware)

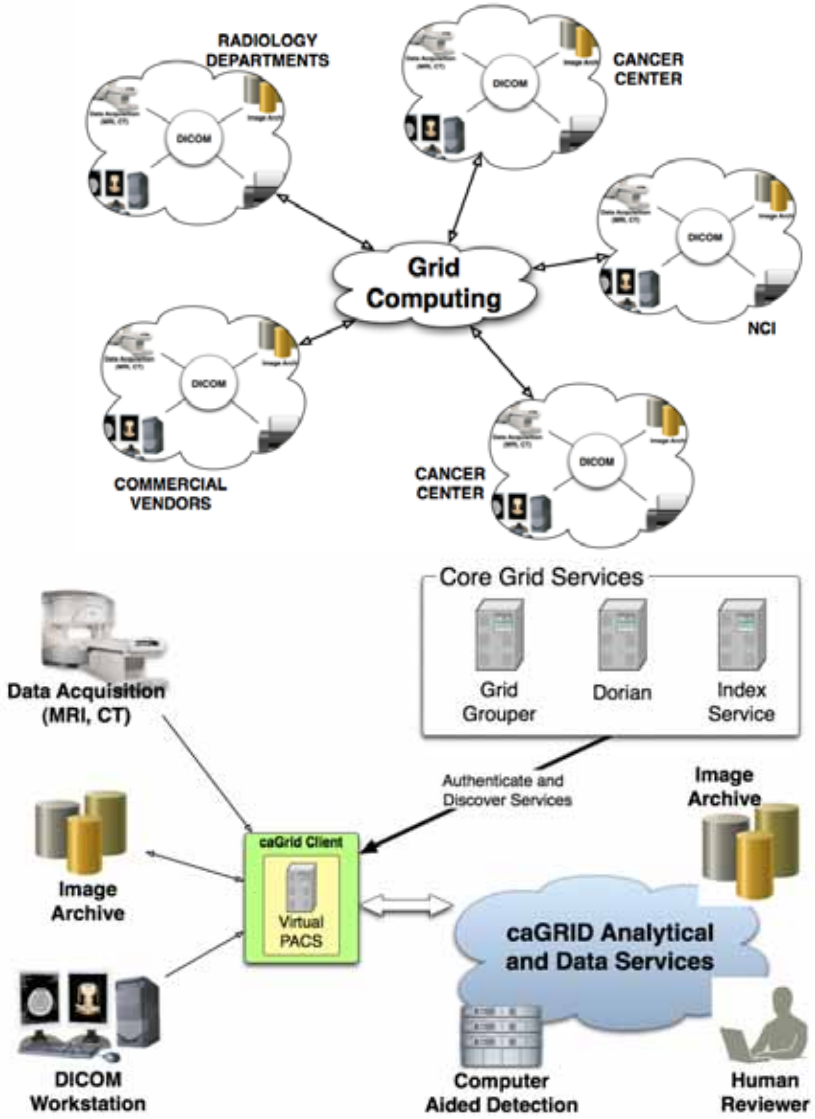


Interoperability Tool: Introduce

- **A framework which enables fast and easy creation of strongly typed and highly interoperable grid services**
- **Provides a powerful extension system wherein specific functionality can be added to the service or service editing process**
  - Support for caDSR, GME, caGrid metadata, Data Services, and caGrid authorization services are all added this way
- **Abstracts all the details of the grid from the developer, allowing them to focus on the business logic being exposed**
- **Provides a graphical environment**

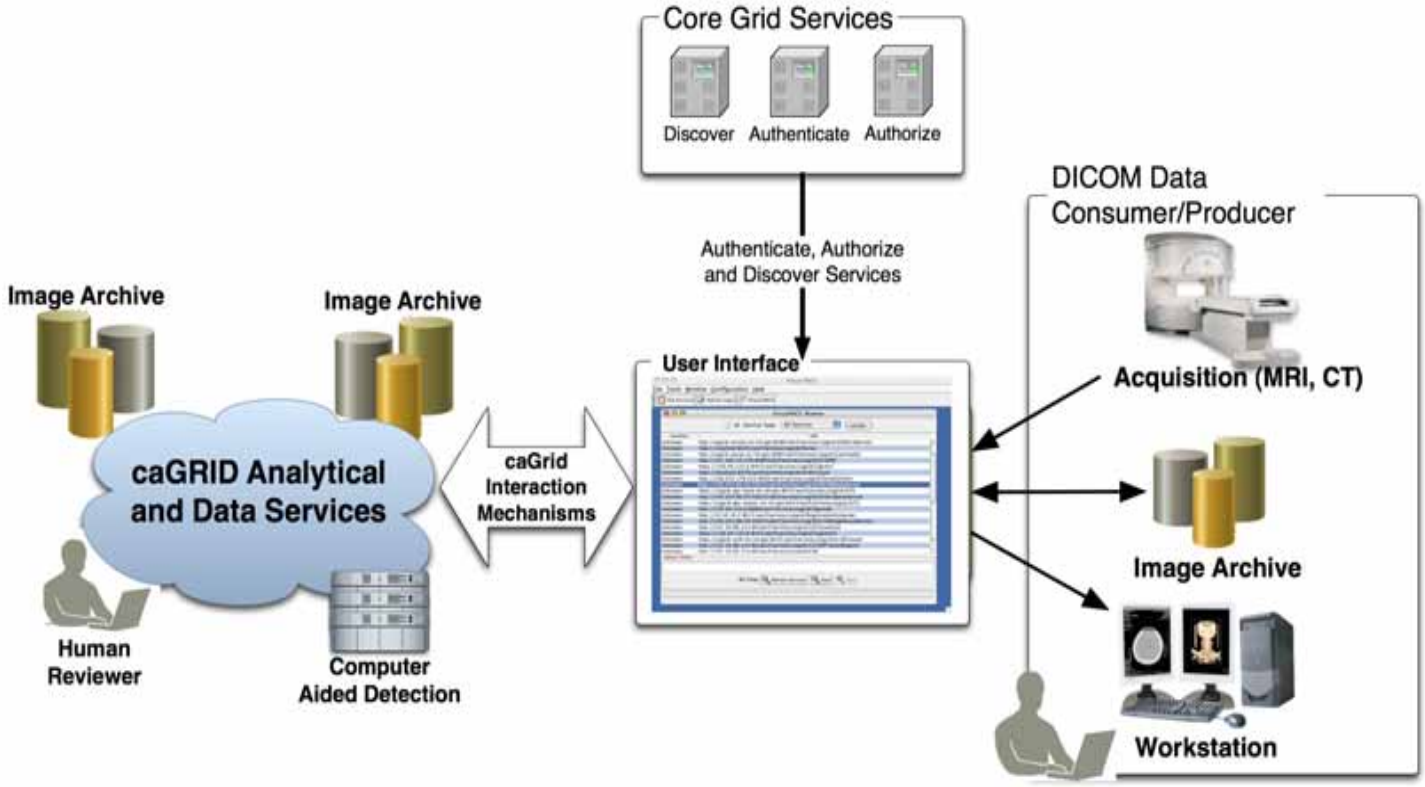
caGrid In vivo Imaging Middleware Extensions

- **Interoperability Library**
  - Translate between DICOM and caBIG data models, and DICOM QR and caBIG query language
- **DICOM Data Service**
  - Exposes existing DICOM QR aware data resources (PACS, etc) as caGrid compliant service
- **VirtualPACS**
  - Allows DICOM-aware clients (review workstation, etc) to access DICOM caGrid data services over the grid
- **caGrid-based security for data transport, authentication, and authorization**
- **Dynamic deployment of services (joint with Globus)**

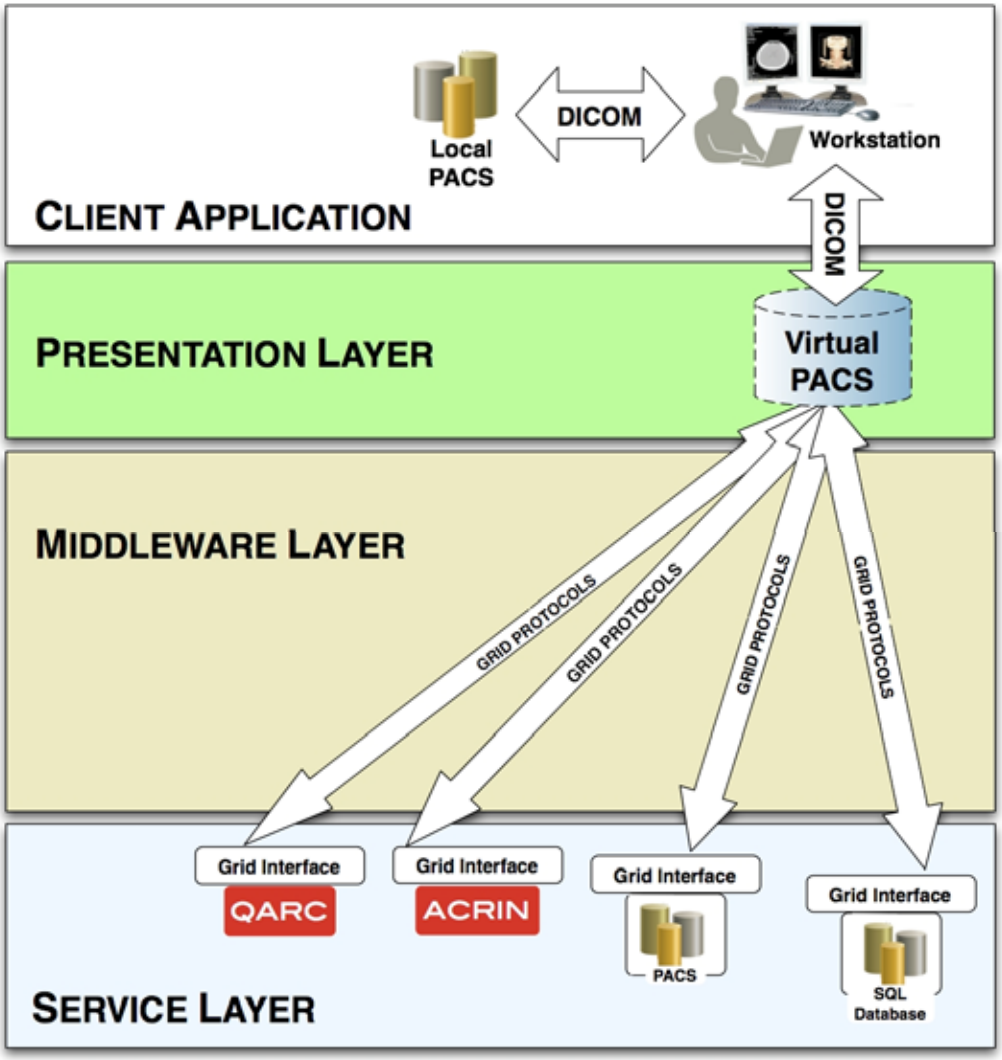


System Architecture - Virtual PACS

Present a PACS interface to analytical and data sources on the grid.
Use your own DICOM Workstation
Virtual PACS federates services on the Grid using caGrid



Three Tier Virtual PACS Design

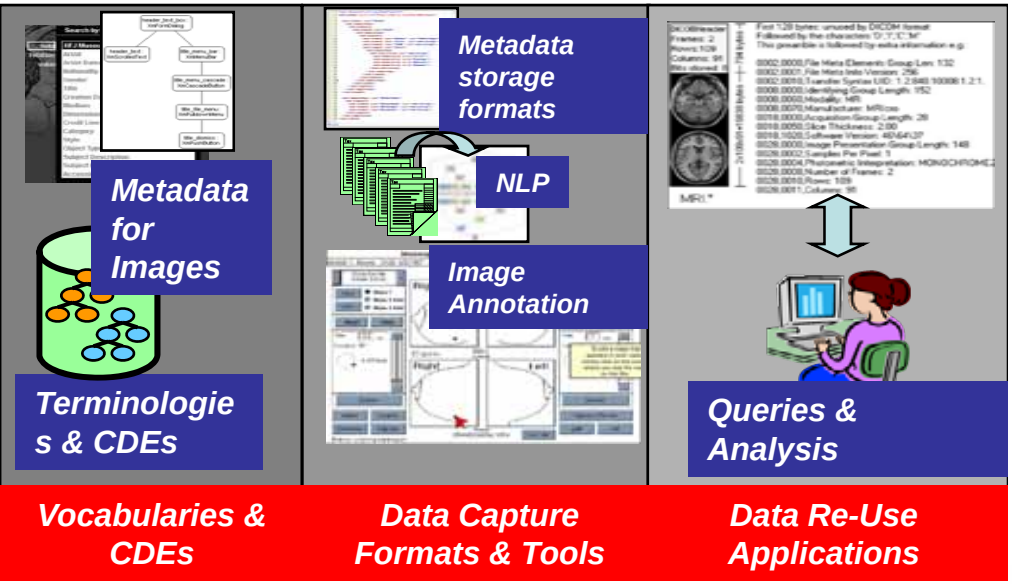


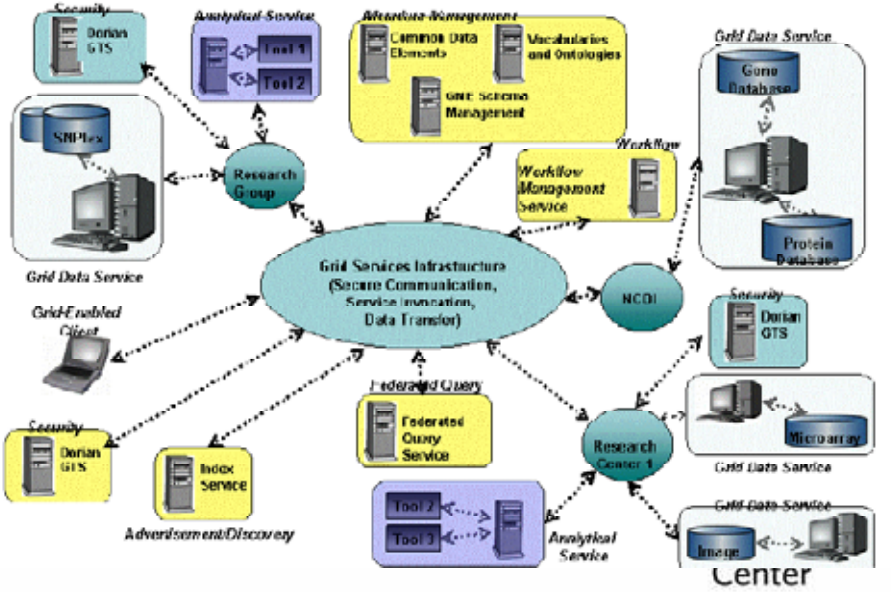
Roadmap

- **Middleware**
- **Images, Cooperative Research Groups and the Grid**

caBIG™ In Vivo Imaging Workspace

- Project 1: Middleware
- Project 2: XIP: Extensible Imaging Platform
- Project 3: AIM: Annotation and Image Markup
- Project 4: Vocabulary



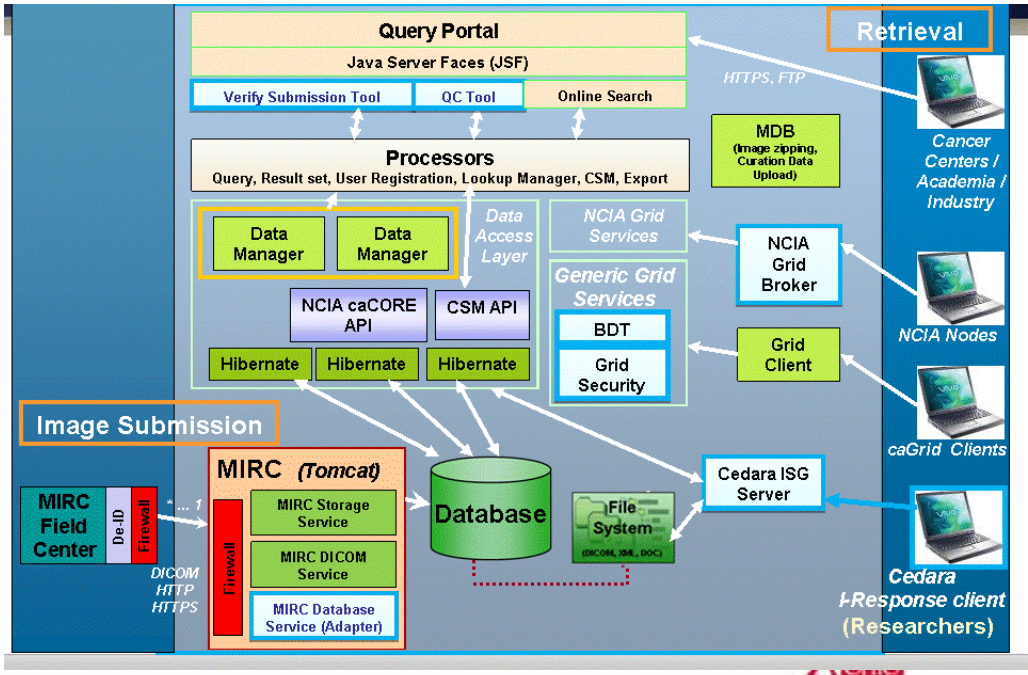


Goals of Imaging Related Cooperative Group Efforts

- ATC, QARC, ITC, CALGB, ACRIN, RTOG, NTROI, Children's Oncology Group
- Maximize reproducibility in staging, grading of cancer
- Maximize reproducibility in radiation treatment planning and delivery
- Accurate characterization of disease to target treatment
- Maximize anatomic and functional precision in targeting tumors
- Imaging as a biomarker

National Cancer Imaging Archive

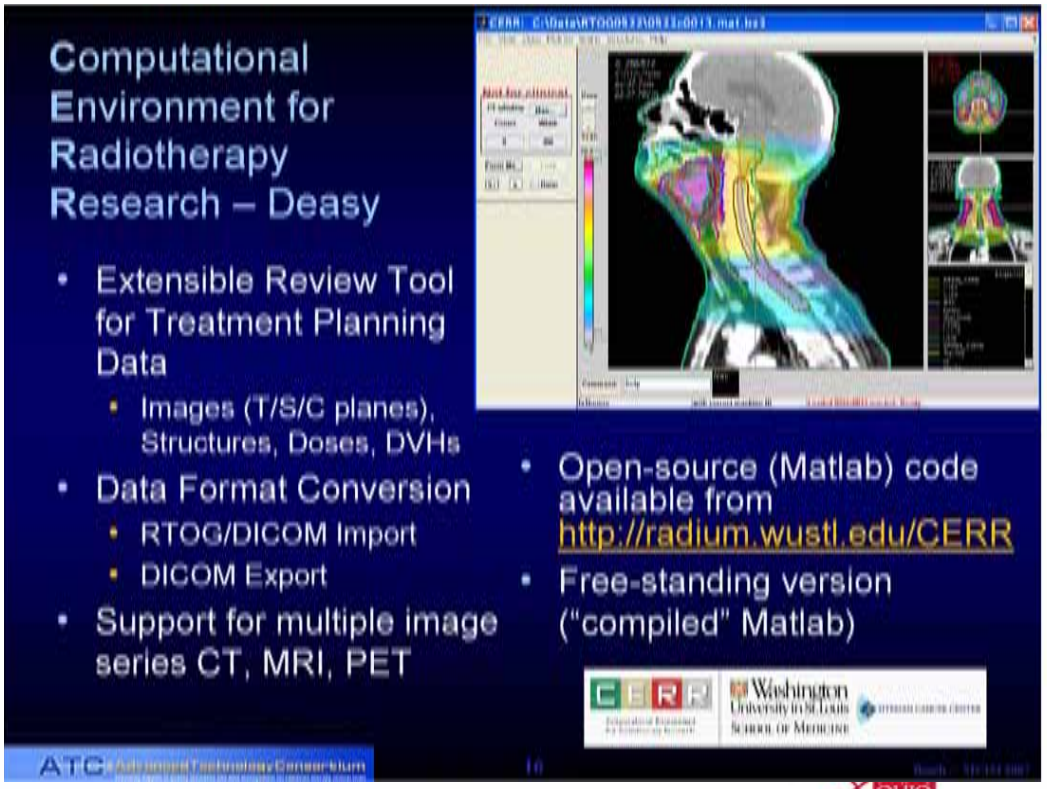
- **Searchable repository of *in vivo* cancer images in DICOM format**
- **Archive currently contains over 2 million DICOM images plus annotation files and curation data**
- **Data Access**
  - Web portal available for accessing repository
  - caGrid access
  - caCORE generated toolkit API
  - Cedara visualization and markup
  - Node-to-Node communication
- **Online searching for cases/images based on criteria such as Image Modality, Slice Thickness, Collection, curation data, etc.**
- **Users can download DICOM data bundled with annotation files and curation data.**



Medical Center

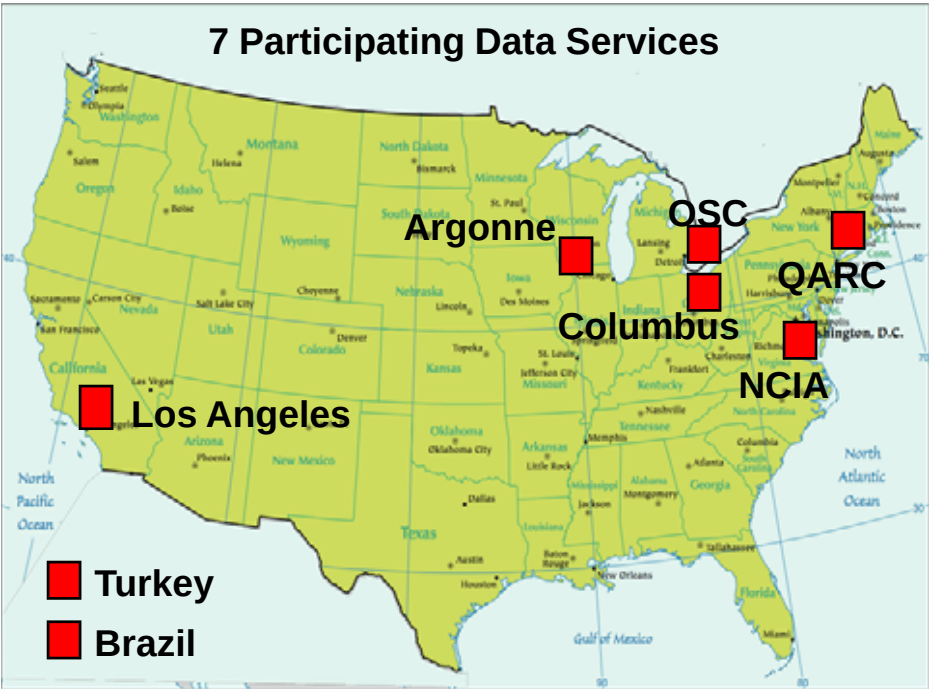
caGrid Enabled CERR

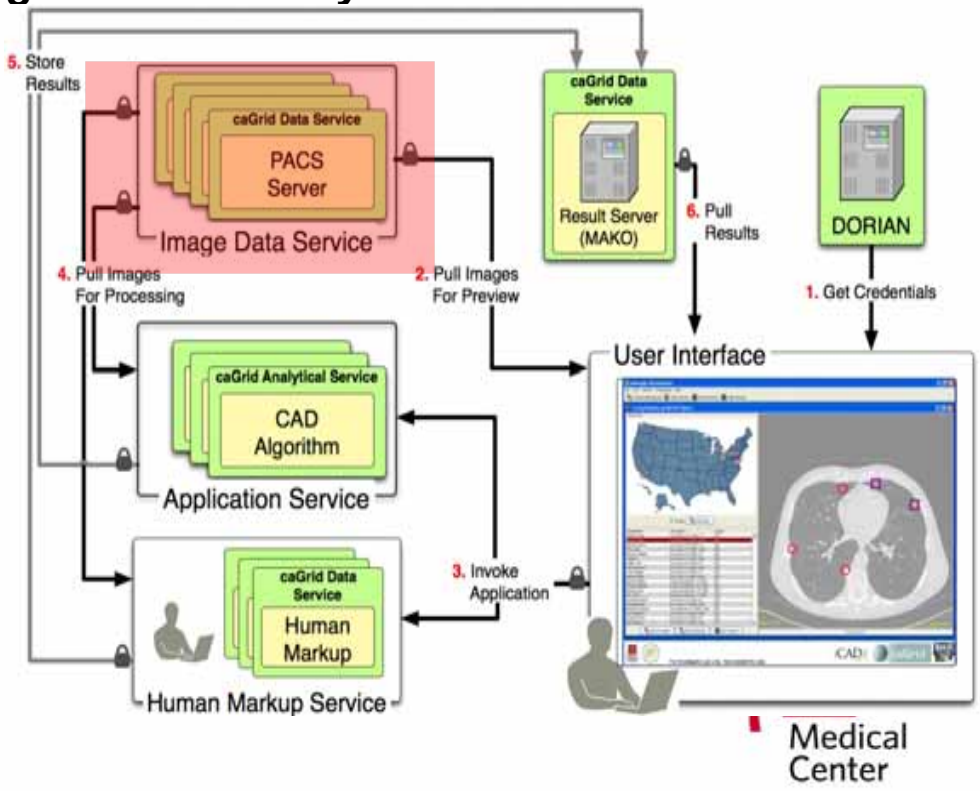
- **Access images, store review results at ATC, QARC, ACRIN, CALGB image archives**
- **Matlab code can run locally or on remote clusters**
- **Markup annotation of images via caBIG AIM standard**



Central Review/CAD Demonstration Project (joint with U Maryland)

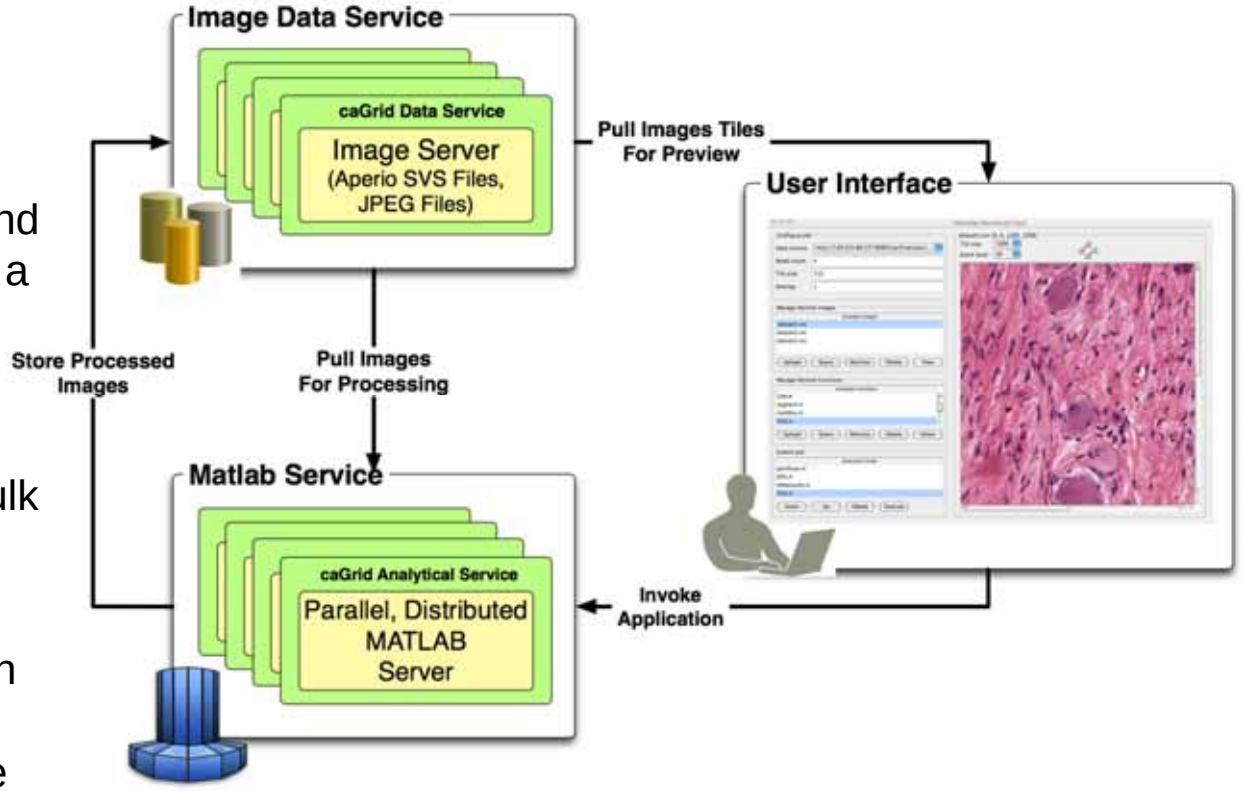
- **Expose data in PACS servers as caGrid Data Service**
- **Participating sites include: Columbus, Chicago, LA, NCIA-Washington DC, QARC Providence, Brazil and Turkey**
- **Set stage for workspace effort to support ACRIN/QARC/CALGB/ATC coordinated review**
- **OSU CALGB Core implemented Grid Image PET/CT analysis/review**



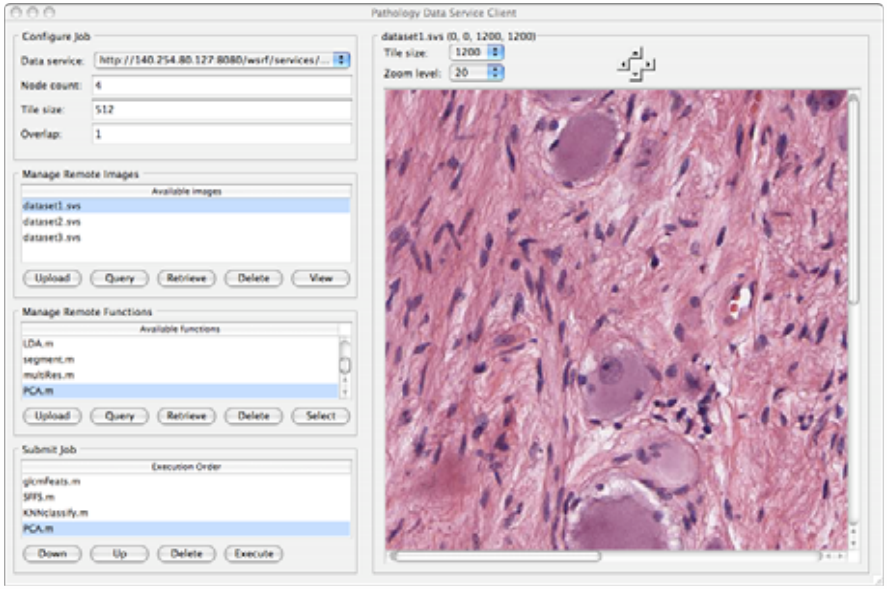


caMicroscope Architecture

- Image Data Services that store images in Aperio SVS format
- Analytical service that pulls images from the image server and executes MATLAB programs on a cluster
- All services are caGrid 1.0, use gridFTP for high performance bulk data transfer
- The Client allows multi-resolution browsing of the images, invoke MATLAB programs and examine the MATLAB output images



Pathology Coordinated Review



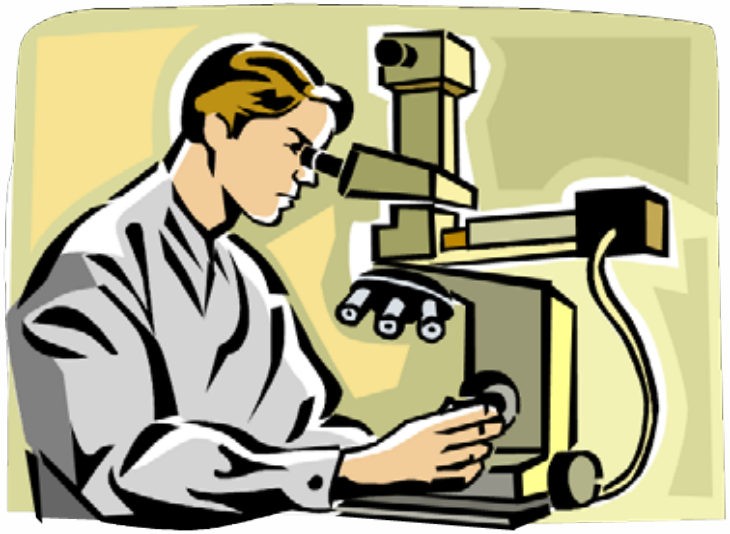
- **Multiheaded Microscope**
- **caMicroscope**

Medical Center

Coordinated Human/Algorithm Pathology Central Review

caBIG cancer Biomedical Informatics Grid™

- Large inter- and intra-reader variability in pathology diagnosis/prognosis
  - 20% variation between central and institutional readers for *neuroblastoma* prognosis [1]
  - Agreement among experts for *follicular lymphoma* grading 61% to 73% [2]
  - Agreement among experts for the identification of *cutaneous T-cell lymphoma* 0.283 – 0.562 (Kappa) [3]



[1] L. A. Teot, R. S. A. Khayat, S. Qualman, G. Reaman, and D. Parham, "The Problems and Promise of Central Pathology Review: Development of a Standardized Procedure for the Children's Oncology Group", Pediatric and Developmental Pathology, **10**, 199–207 (2007)

[2] F. Dick, S. VanLier and P. Banks, et al, "Use of the working formulation for non-Hodgkin's lymphoma in epidemiological studies: agreement between reported diagnoses and a panel of experienced pathologists," Journal of National Cancer Institue, **78**, 1137-1144 (1987)

[3] M. Santucci, A. Biggeri, A. C. Feller, and G. Burg, "Accuracy, concordance, and reproducibility of histologic diagnosis in cutaneous T-cell lymphoma: an EORTC Cutaneous Lymphoma Project Group Study. European Organization for Research and Treatment of Cancer," *Arch Dermatol,* vol. 136, pp. 497-502, Apr 2000.

Ohio State University Medical Center

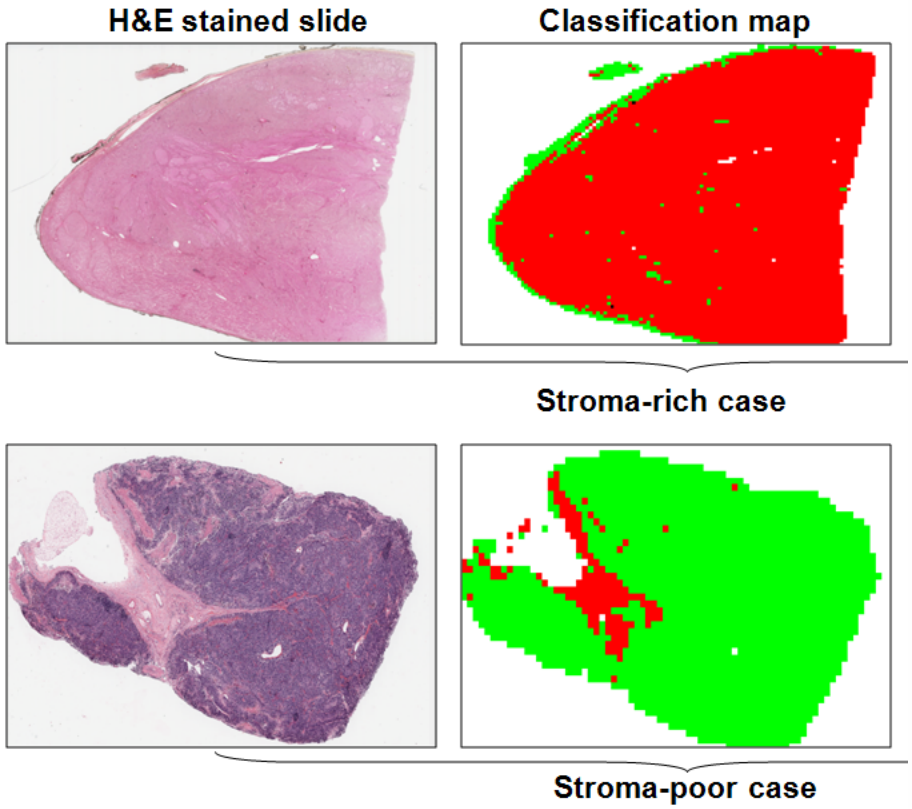
Computer-assisted Histopathology

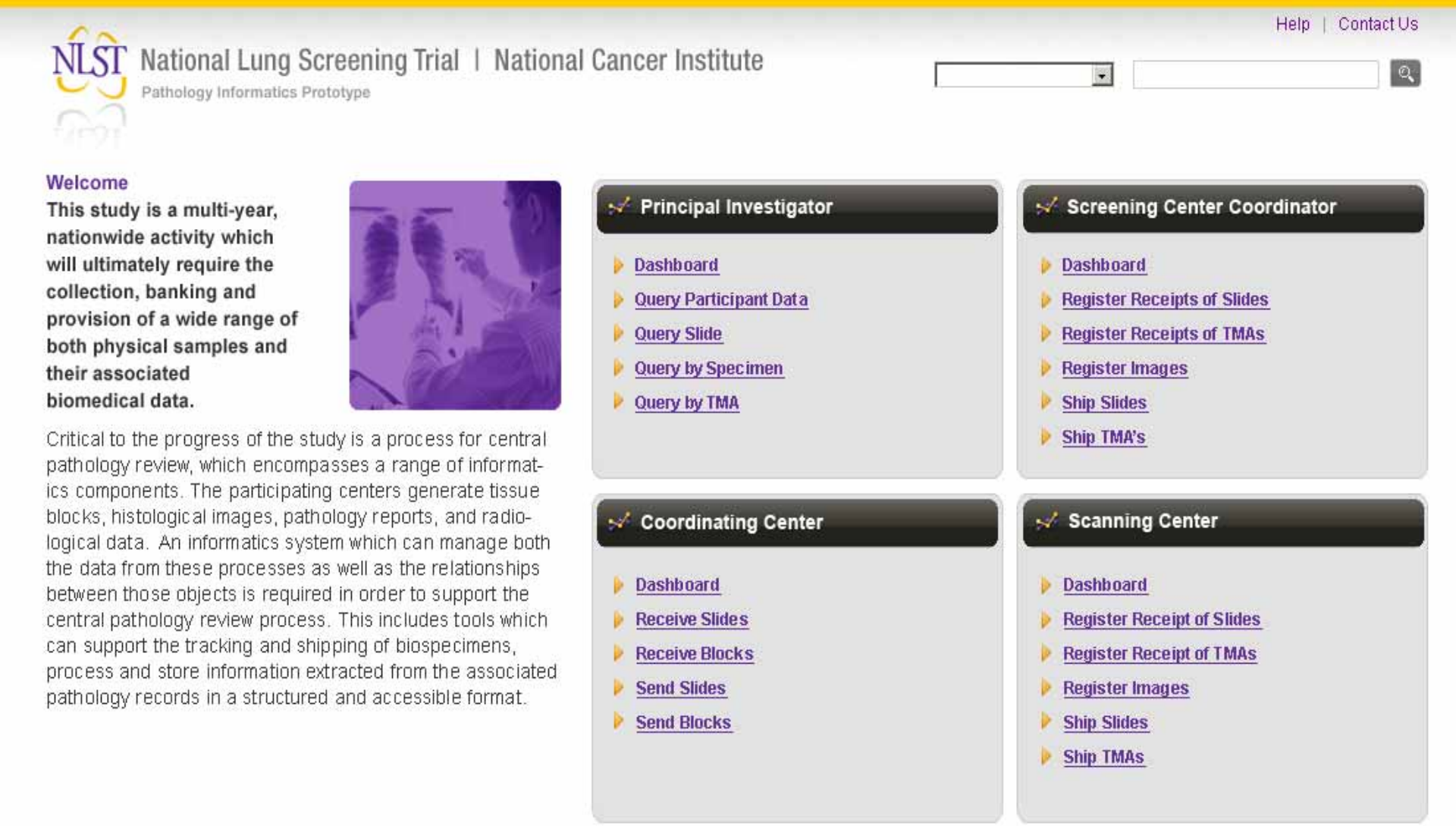
- Analyze images by computer
- Analyze the whole tissue, several slides
- Provide quantitative information to the pathologist
- Reduce inter- and intra-reader variability

Medical Center

Computerized Analysis of Slides

- **Neuroblastoma:**
  - Stroma: 88.4% accuracy
  - Differentiation: 87.9% accuracy
- **Follicular Lymphoma:**
  - Grade III identification:
    - 98.9% sensitivity
    - 98.7% specificity
- **Breast Cancer (Foran)**
  - Classification from TMA
    - 86.02% accuracy
- **Classification of hematologic malignancies (Foran)**
  - Case-based classification 89% accuracy





- **Track specimens using caTissue**
- **Support coordinated review of digitized slides using NCIA, caMicroscope**

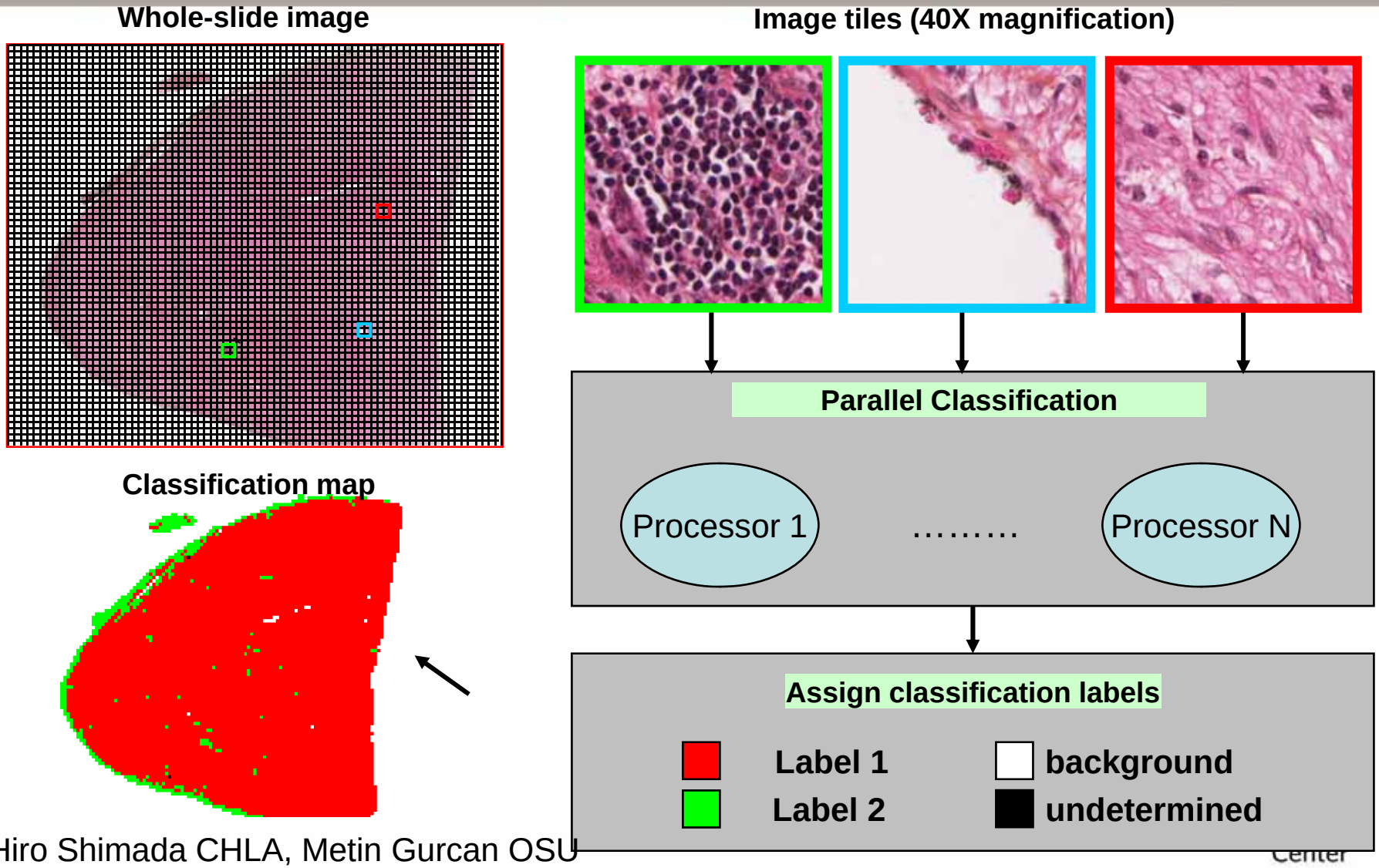
Portal

Multi-institutional, Grid-enabled Digital Pathology Testbed

Lead Investigators: Foran and Saltz

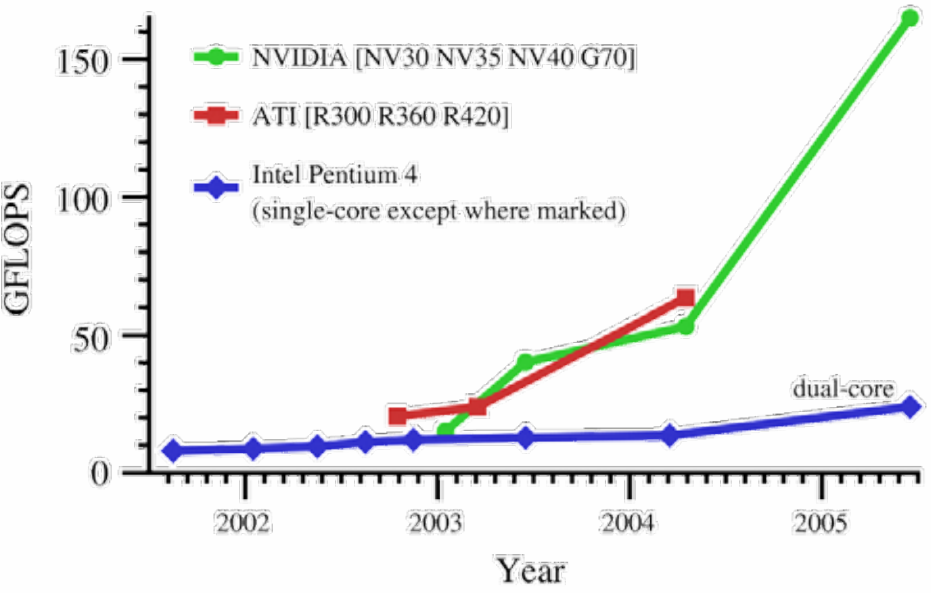
David J. Foran, Ph.D.

caMicroscope parallel processing
Children's Oncology Group - Neuroblastoma



Hiro Shimada CHLA, Metin Gurcan OSU

Use of GPU resources

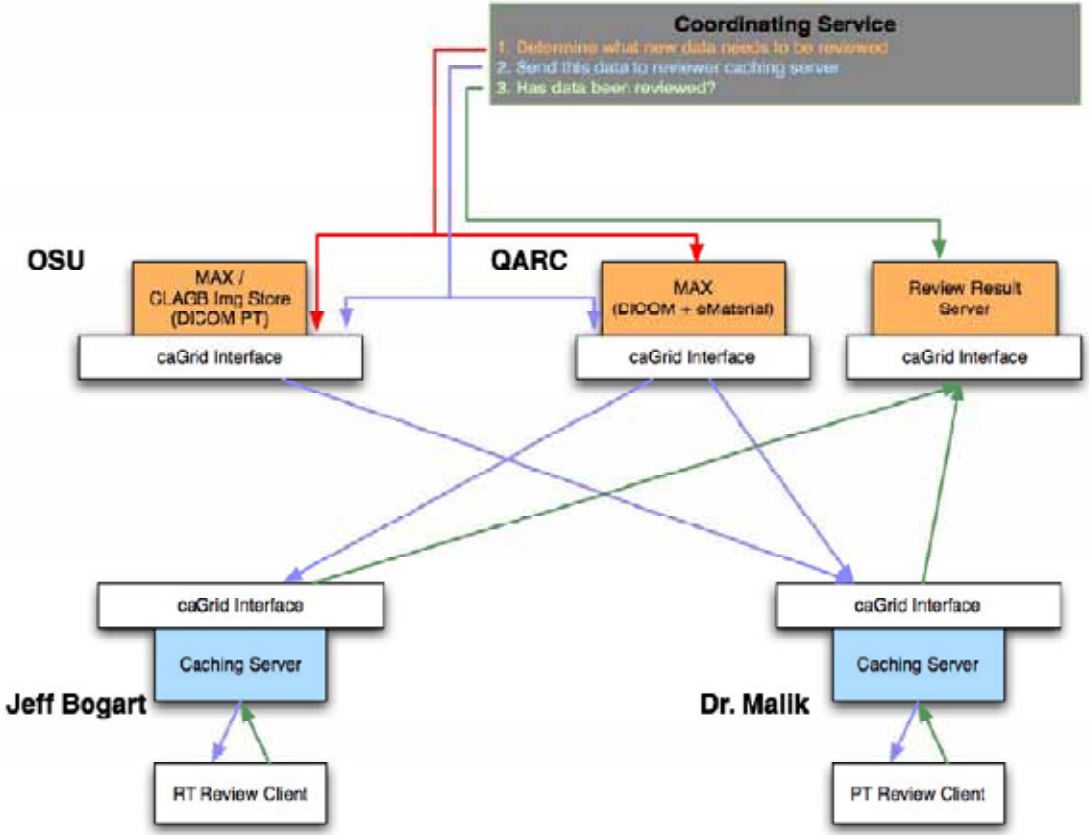


- **GPUs:**
  - Speed increasing at cubed-Moore's law!
  - Ubiquitous and inexpensive
  - Functional units for specific graphics-based operations (vertex & pixel shaders)
  - Small memory but raw computational power
  - Memory bandwidth & clock provides superior performance

	CPU	GPU
Processor clock	**2.13 GHz**	**575 MHz**
Raw computational power	**10 GFLOPS**	**520 GFLOPS**
Memory bus width	**64 bits**	**384 bits**
Memory clock	**2x333 MHz**	**2x900 MHz**
Memory bandwidth	**10.8 GB/s**	**86.4 GB/s**
Memory size and type	**2 Gb DDR2**	**768 Mb GDDR3**

CTTI

Pilot study to evaluate the application of caGrid in vivo imaging technologies in imaging based NCI clinical trials



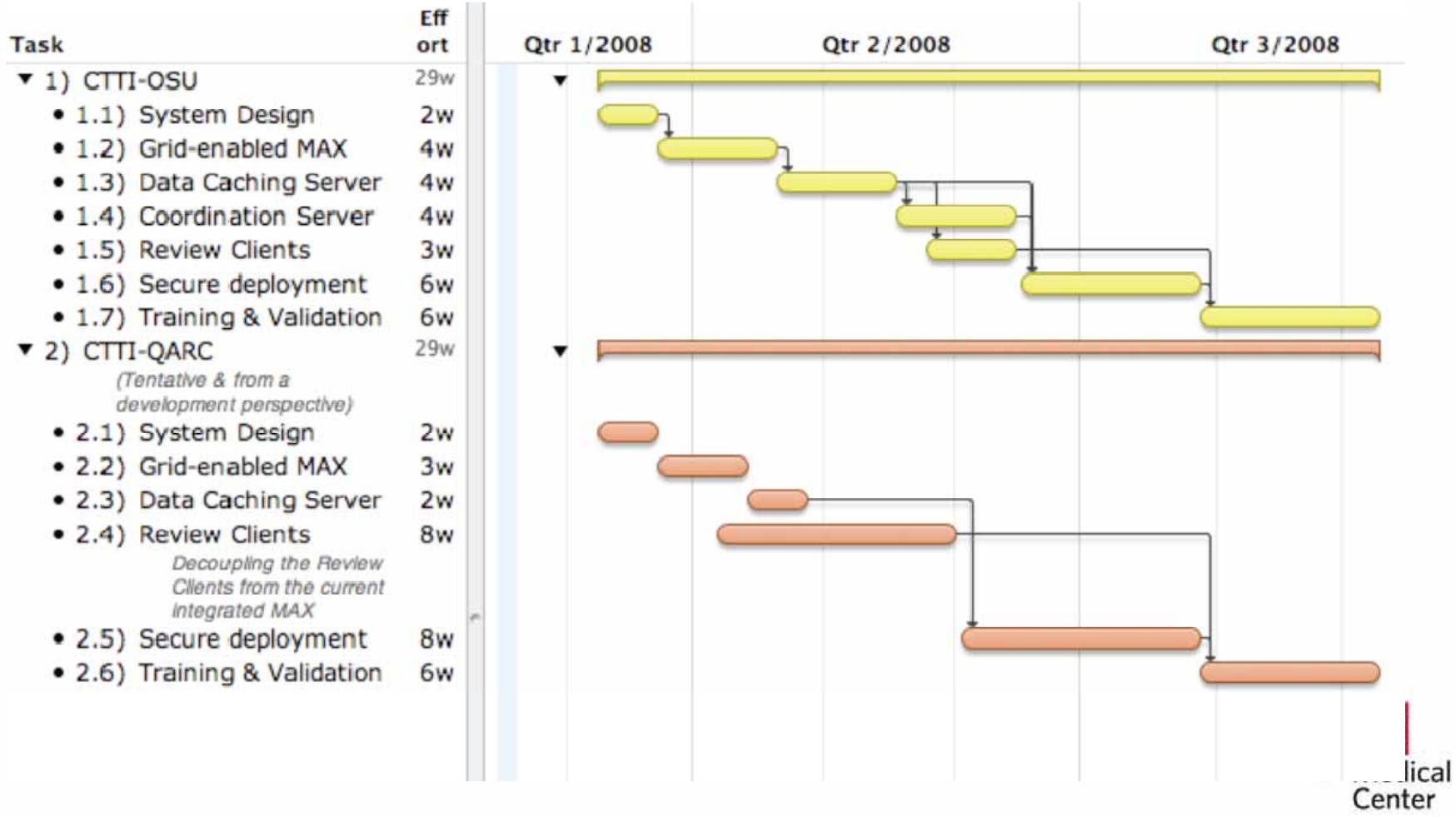
Medical Center

- **Develop and deploy a data and security infrastructure and review clients for use in the central review process for CALGB 80302 protocol.**
- **Develop a grid enabled data coordination service to manage the data and track the review process.**
  - Monitors: 1) What new data has been submitted to QARC and is marked as ready for review, 2) Push this data to the remote reviewers caching server, 3) Monitor the result server to see if and when reviews are completed.
- **Develop a data caching server, hosted at each of the review sites. Pushes data to the caching server.**
  - Review client uses data copies in caching server
  - Caching server pushes review results to the centralized result server at QARC. If the caching server does not have objects requested by a reviewer, it will have the ability to request this data from the centralized server at QARC.
- **Grid enablement of MAX-database to allow the various objects needed in the review process to be accessible across the grid.**
- **Review clients used in this study will be developed in two phases.**
- **In Phase 1, MAX-PT clients and CERR will be used for review. Since both these clients have different methods of query/retrieval/submission, caching servers will be customized to support these clients.**
- **In a subsequent Phase 2, a review client will be developed using XIP. Other requirements for this XIP based review client will evolve from the reviewer experience in Phase 1.**
- **A Data models to describe the PT & RT review. This model could be an 80302 specialization of the AIM model.**

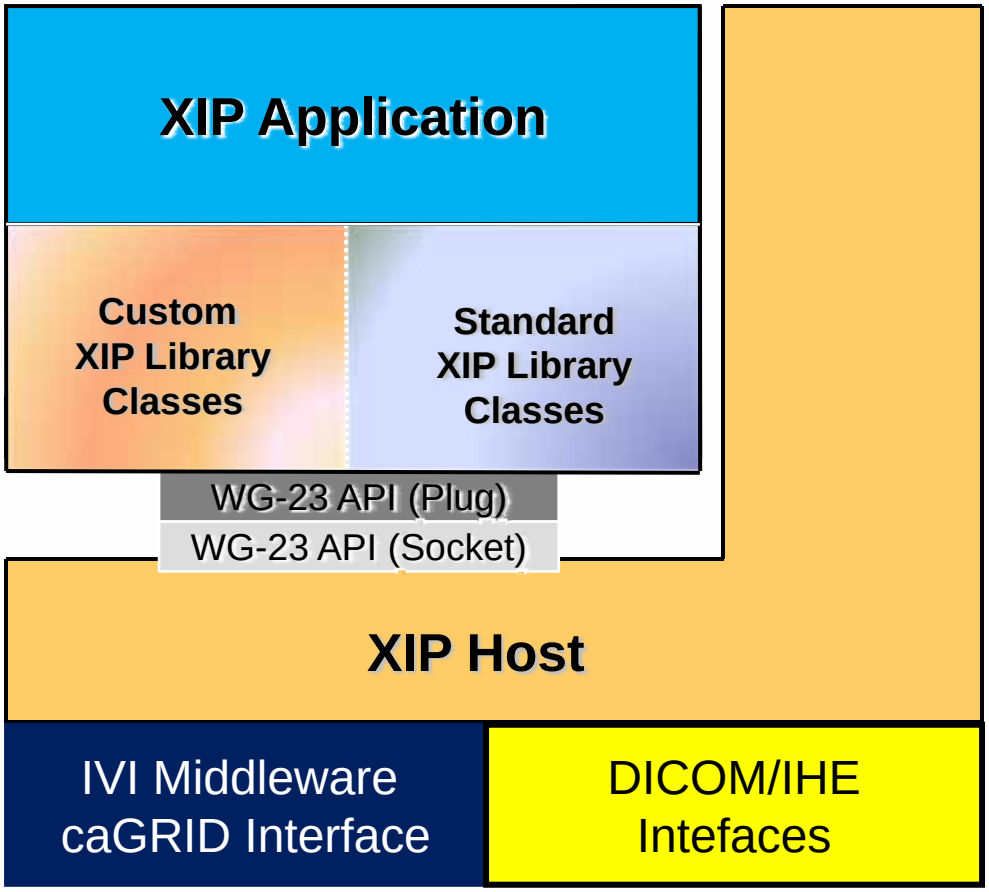
Core Infrastructure Components

- **Data Coordination Service - Manages the review workflow.**
  - Rule-based engine that orchestrates all data transfers needed for a specific stage of the review cycle. For example when a review has to take place this service will create the work-orders for what data is needed and subsequently provide this data to the raters. Similarly when a review has been completed, it will create a work-order for the adjudicator who can then harmonize the evaluations of the different reviewers.
  - Data coordination service will be needed by multiple Imaging workspace projects including the ACRIN 6675 support project and the NCIA curation and visualization client project. Both these projects include use-cases involving clients that retrieve data from grid resources in response to work-orders created for the different user roles. Data Coordination Service will be generalized and contributed back to the core imaging middleware stack.
- **Data Caching Server – supports data push from data curation site to reviewer sites.**
  - Data services are designed to serve data in a traditional client-server architecture where the clients requests and pulls the data from the server. Performance considerations or security constraints are sometimes better satisfied by a push based mechanism.
  - In some cases, institutional security restrictions do not permit remote grid clients to pull data from grid-services. In other cases slow internet connections make it advantageous to push data to a remote cache. The VIEW consortium is one such motivating use-case where such a push strategy of a data caching server could be used to disseminate, QARC data ACRIN and to Ohio State.

Timeline (time shift as necessary to account for start date)



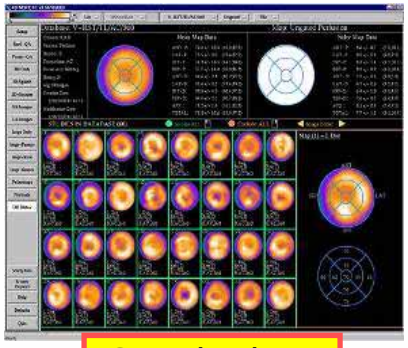
XIP Architecture



DICOM WG23 Plug-in Framework

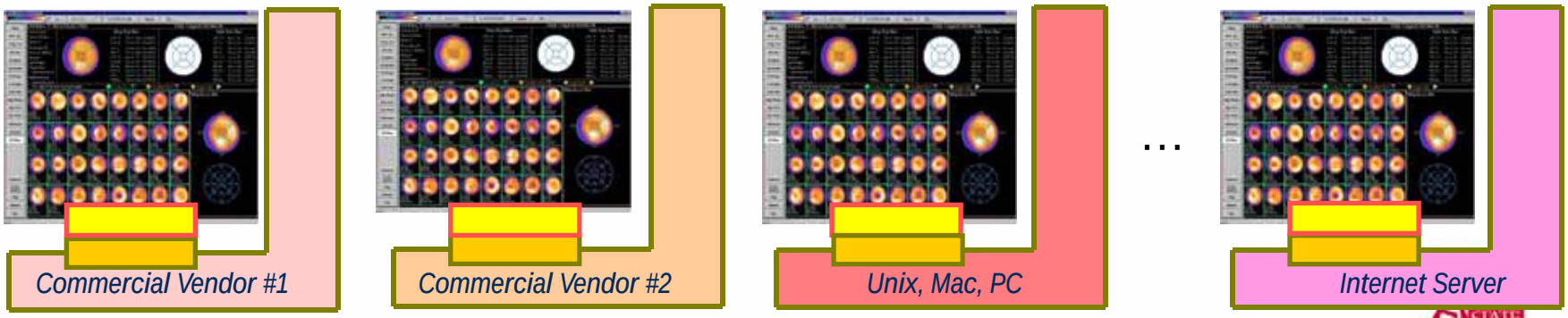
- WG-23 addresses clinical integration and vendor inter-operability by defining standardized “plugs” and “sockets” (APIs)

- caBIG XIP addresses an open-architecture, integrated development environment for rapid prototyping & collaboration based on WG 23 APIs. It will be a reference implementation of those APIs



XIP developed Application

Standard API



← Clinical →

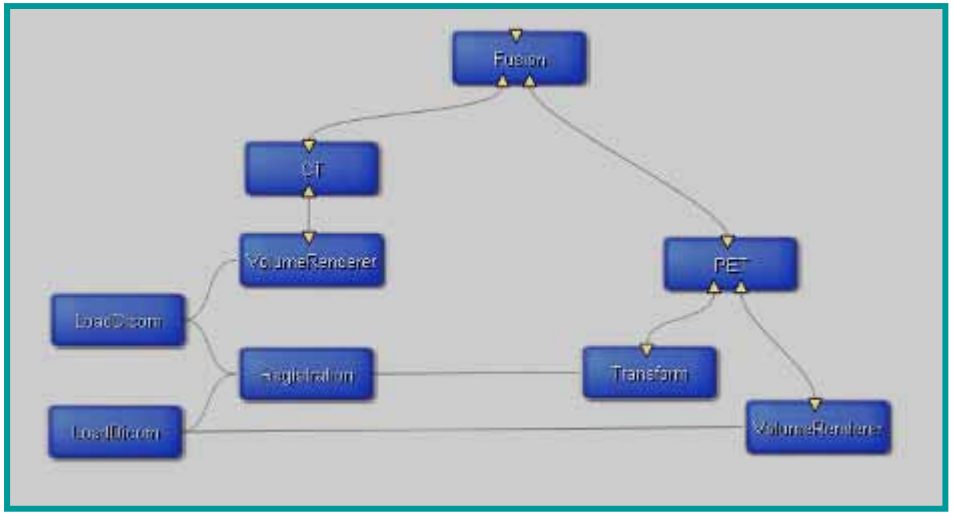
← Prototype & Collaboration →

XIP Libraries

- **Sets of Open Inventor ® objects that XIP applications may utilize**
  - Stock Open Inventor® objects for 3D graphics, modeling, UI, etc.
  - Standard XIP objects for retrieving data from, and storing data to XIP or DICOM WG-23 Hosts via the DICOM WG-23 interfaces
  - XIP objects that extend Open Inventor® objects with functions optimized for medical imaging applications, including MPR, image markup, annotations, etc.
  - XIP objects for accessing GPU hardware for volume visualization, filtering, etc.
  - VTK classes wrapped into Open Inventor® objects to support visualization
  - ITK classes wrapped into Open Inventor® objects to support image processing
  - Custom objects supplied by developers to extend XIP
- **Many XIP Libraries are host-independent, but if dependent on host services:**
  - XIP libraries have internal dynamic ‘adapter’ libraries to allow for easy switching between XIP hosts or platforms (i.e. simply swap the dynamic library)
  - The XIP Application always sees the same Open Inventor® objects regardless of host
- **XIP libraries may be auto-generated from existing class libraries (e.g. ITK, VTK), or may be custom-built from new or existing code**

Open Inventor®

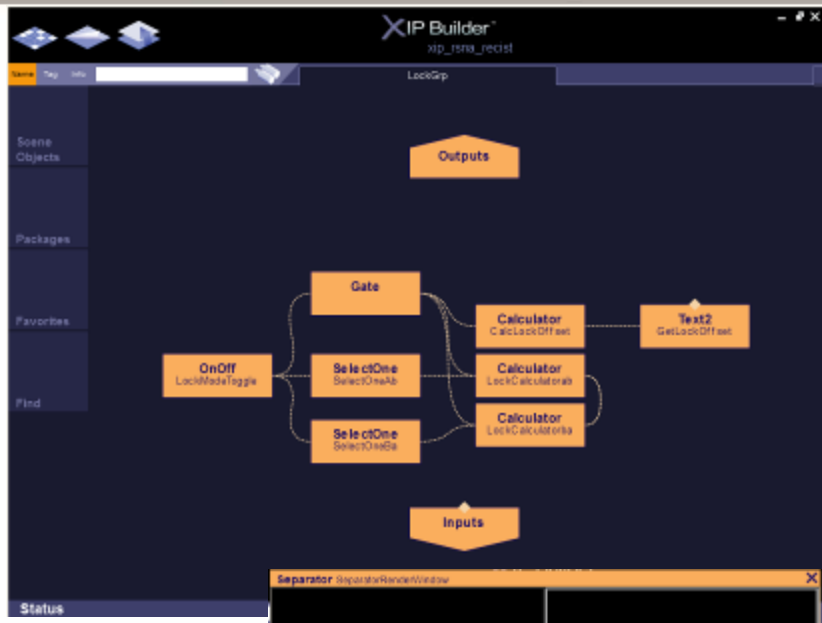
- Open Inventor® is an object-oriented 3D toolkit for interactive graphics programming
- Open Inventor® presents a programming model based on the **Model/View/Controller** design pattern
- Open Inventor® includes both the **Pipeline** (for processing data) and **Scene Graph** (for visualizing data and models) concepts
- **Open Inventor® has a simple extension mechanism**, making it easy to add new libraries (e.g. VTK, ITK) and functions
- Open Inventor® modules represent **Engines, Nodes and Manipulators**

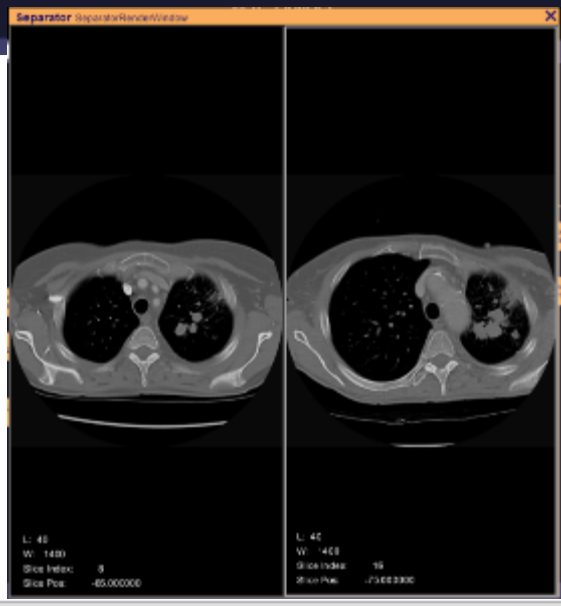


Medical Center

XIP Applications

- **Applications formed from Interconnected Open Inventor ® engines, nodes, and manipulators from XIP Libraries**
  - **Engines** enable the creation of processing pipelines and animation
  - **Nodes** support the concept of scene graphs, which are hierarchical structures of objects describing what needs to be visualized in 2D/3D, fed by the engines
  - **Manipulators** handle input devices, measurements and coordinate transforms in response to user interaction via a simple event model
- **Focused on the processing logic, not the infrastructure or data format**
- **Skinnable GUI engine for different look-and-feel on different platforms**





XIP Application Builder

XIP Development Process

Medical Imaging Workstation

Distribute

XIP Application

XIP Modules

Host Independent

ITK | VTK | XIP LIB | ...

XIP Host Adapter

WG 23 | WG 23 | WG 23 | WG 23

XIP Host

(Can be replaced with any DICOM WG23-compatible Host)

Host-Specific Plug-in Libs

DICOM, HL7, & other services per IHE

caGRID Services via Imaging Middlewar

Distribute

Web-based Application

Distribute

Standalone Application

XIP Class Library

Auto Conversion Tool

Thank you

Acknowledgments

- The caGrid team:caGrid 1.0: Scott Oster, Stephen Langella, Shannon Hastings, David Ervin, Ravi Madduri, Tahsin Kurc, Frank Siebenlist, Ian Foster, Krishnakant Shanbhag, Peter Covitz
- OSU Imaging Informatics/HPC team: Berkant Barla Cambazoglu Ph.D., Umit V. Catalyurek Ph.D., Metin N. Gurcan Ph. D. ,Kun Huang Ph.D.,Tony C. Pan M.S., Ashish Sharma Ph.D., Manuel Ujaldon Ph.D. Olcay Sertel Antonio Ruiz, Vijay Kumar
- Jim Purdy, Walter Bosch, Joe Deasy from Advanced Technology Consortium
- Eliot Siegel, Paul Mulhorn, Michael McNitt-Gray, all SMEs and participants in the caBIG in-vivo imaging workspace
- Baris Suzek (Georgetown), Kiran Keshav (Columbia), Rakesh Nagarajan (Wash U), Carl Blake, Anand Basu, John Freymann (NCIA project)