



# **Data Standards and Image Handling at the NCI Center for Bioinformatics**

**Peter A. Covitz, Ph.D., NCICB**

**Fei Xu, Ph.D., NCICB**

**Sharon Settnek, SAIC**

**Cancer Imaging Informatics Workshop  
September 26, 2002**

# Part I. Cancer Data Standards Repository

***Metadata management system for  
NCI Common Data Elements***





# NCI Center for Bioinformatics

- ◆ Established in early 2001 by Kenneth Buetow
- ◆ Folded in several existing programs, launched new ones
- ◆ Mission to bridge basic and clinical research via bio- medical- informatics



# NCICB-supported initiatives

- ◆ Cancer Genome Anatomy Project
  - *ESTs, SAGE, SNPs, Pathways*
- ◆ Molecular Signatures of Cancer
  - *Microarray-based tumor classification*
- ◆ Mouse Models of Human Cancer Consortium
  - *Pathology, microarray*
- ◆ Clinical Trials
  - *Protocols, agents, targets, patient cases*



# caCORE

- ◆ Core Infrastructure for biomedical application development
  - NCI Enterprise Vocabulary Services  
*Thesauri and Ontologies*
  - Bioinformatics Infrastructure Objects  
*UML-modeled middleware and APIs*
  - Cancer Data Standards Repository  
*Metadata*

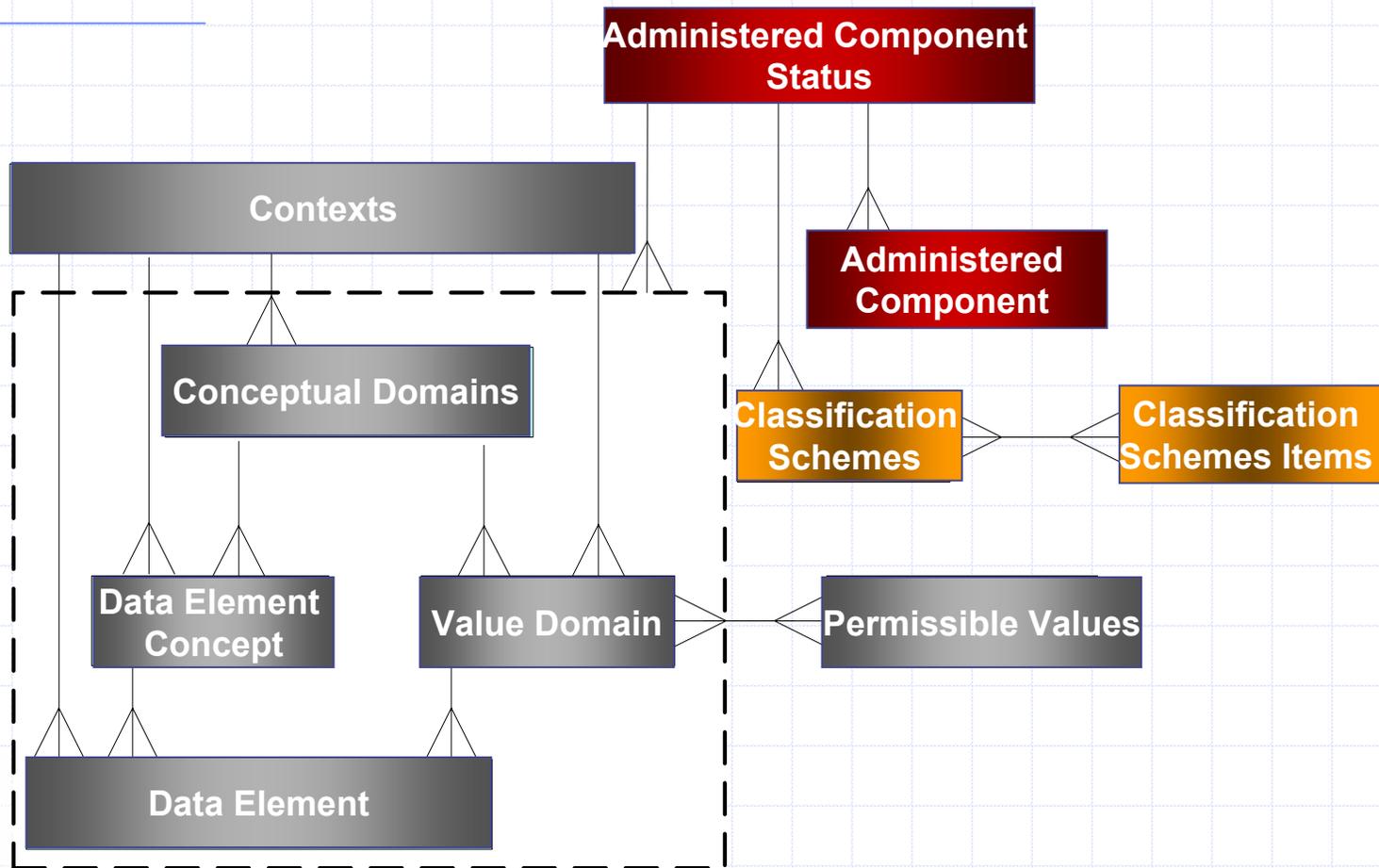


# Cancer Data Standards Repository (caDSR)

- ◆ Registry for Common Data Elements
- ◆ Implemented in Oracle DBMS
- ◆ Logically separate Contexts
- ◆ caDSR Tools:
  - CDE Form Compliance Review
  - CDE Browse/Query/Export
  - CDE Curation
- ◆ caDSR Programming Interfaces:
  - PL/SQL
  - XML

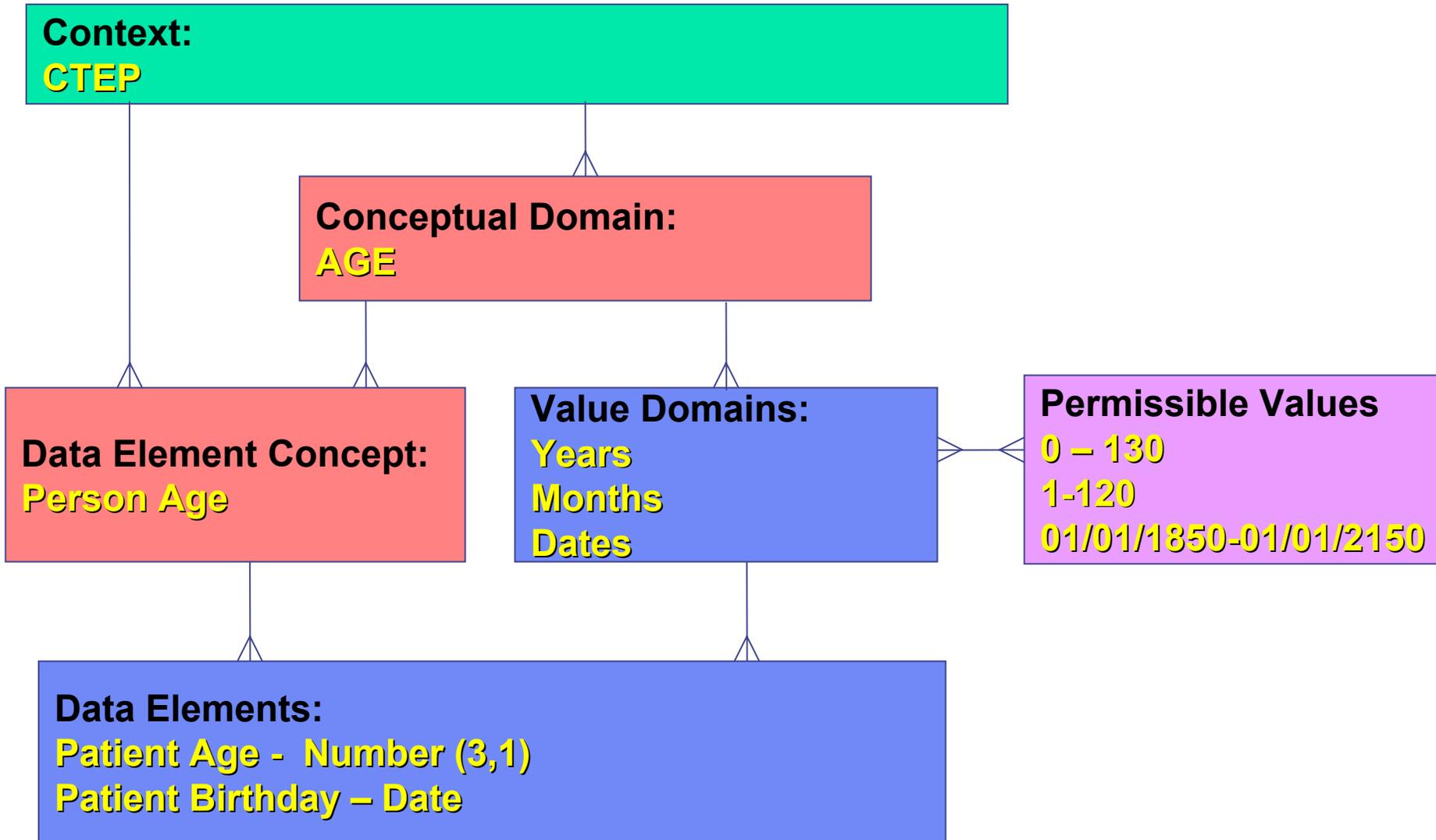


# ISO/IEC 11179 Model





# Example CDEs





# Common Data Element Contexts



**SPOREs**  
Specialized Programs  
of Research Excellence



Cancer Biomarkers Research Group
Early Detection Research Network



NATIONAL CANCER INSTITUTE

**Biomedical Imaging Program**

Division of  
**CANCER PREVENTION**

**Different Curators for each Context**





# CDEs in Production

- ◆ Clinical Trials Support Unit
  - Cooperative Group trials overseen by CTEP must transmit data to the CTSU using CDEs
- ◆ Clinical Data Entry System
  - NCI Center for Cancer Research trials implementing CDEs in the new Oracle Clinical-based system being deployed
- ◆ Early Detection Research Network



# caDSR Next Steps

- ◆ Tool and Interface Completion
- ◆ Additional CDE Domains
  - Pediatrics
  - Phase I/II Trials
  - Genomics
- ◆ Harmonization of CDEs...
  - Across Contexts, if it makes sense to do so
  - With Public Standards (HL7, LOINC, etc.)



# Info and mailing lists

◆ caDSR web site

<http://ncicb.nci.nih.gov/core/caDSR>

◆ caDSR users list

[http://list.nih.gov/archives/sbr\\_users.html](http://list.nih.gov/archives/sbr_users.html)



# Acknowledgements

**NATIONAL  
CANCER  
INSTITUTE**

- Kenneth Buetow
- Denise Warzel
- Beverly Meadows
- Jeff Abrams

**ORACLE®**

- Christo Andonyadis
- Jasur Ishmukhamedov
- Ram Chilukuri
- Prerna Aggarwal



**ScenPro, Inc.**

- Bill McCurry

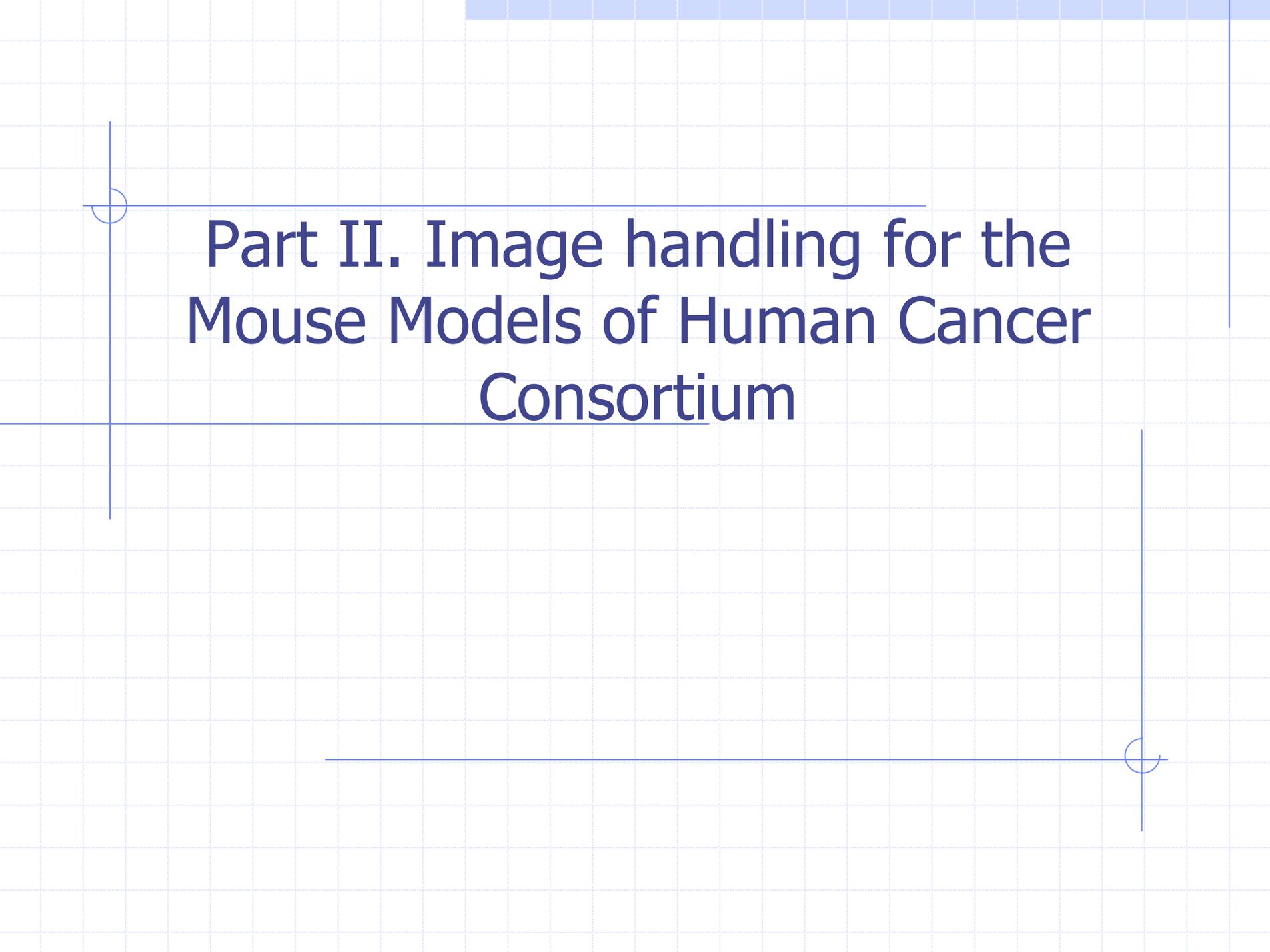


**The EMMES Corporation®**

- Claudine Valmonte
- Pam West
- Mary Supley

**NATIONAL  
CANCER  
INSTITUTE**





# Part II. Image handling for the Mouse Models of Human Cancer Consortium



# Image Portal

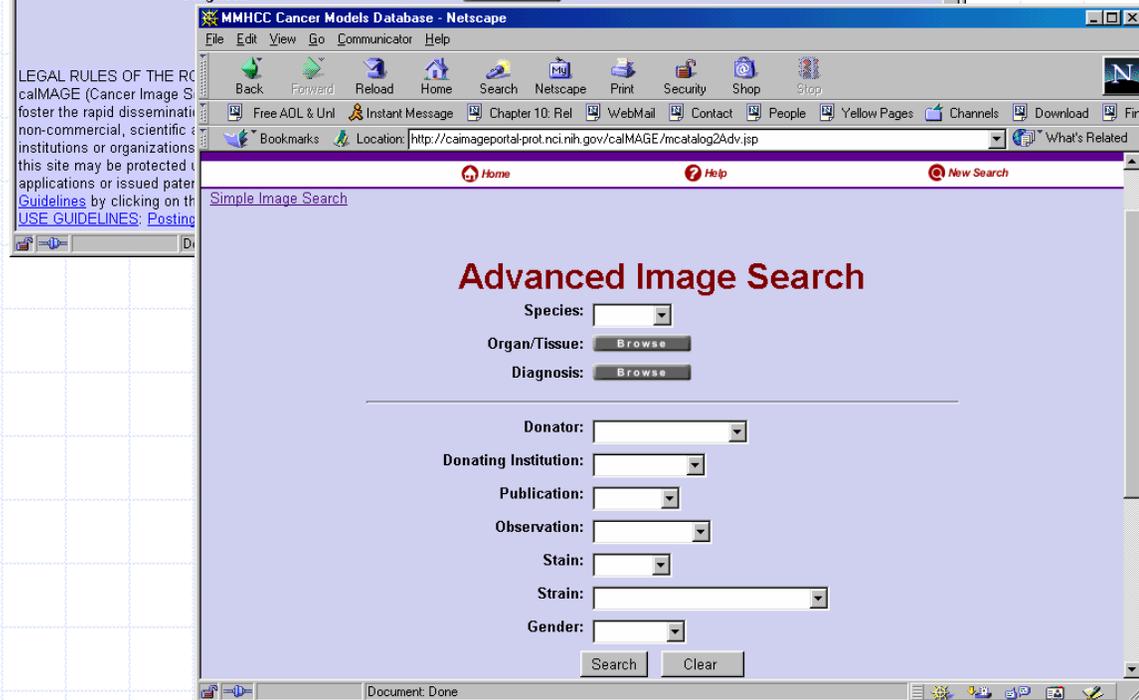
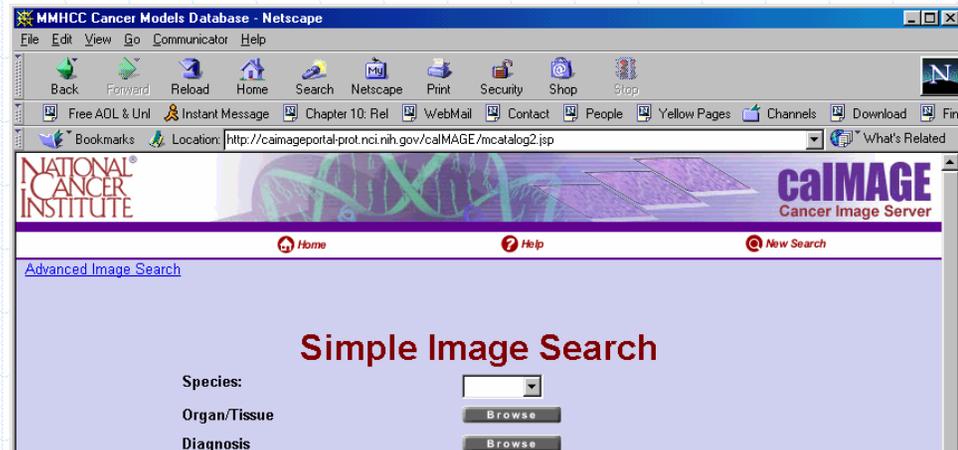
- The NCICB has developed a cancer image portal to allow researchers to search for mouse and human images and annotations
- Human and mouse images and annotations were provided by the MMHCC

The screenshot shows the caIMAGE Cancer Image Server website. The browser window title is "NCI - Netscape". The address bar shows the URL "http://cancerimages.nci.nih.gov/caIMAGE/index.jsp". The website header includes the "NATIONAL CANCER INSTITUTE" logo and the "caIMAGE Cancer Image Server" title. The main content area features three columns of links: "Search Images" (with a circular image of a DNA helix), "Submit Images" (with a circular image of a DNA helix), and "Admin" (with a circular image of a laptop). To the right of these links are four red arrows pointing right, and a "Disclaimer" link. Below the links is a "LEGAL RULES OF THE ROAD" section with a small text box. At the bottom, there is a footer with the text "Please send comments and suggestions to emice@pop.nci.nih.gov" and links for "Privacy Notice" and "Accessibility Information".



# Image Search

- ◆ Researchers can perform basic and advanced image searches
- ◆ Researchers can search images based on common vocabulary





# Image Vocabulary

- ◆ The Enterprise Vocabulary System (EVS) is a set of services and resources that address NCI's needs for controlled vocabulary
- ◆ The Image Server utilizes EVS to access controlled vocabulary for
  - Organ/Tissue
  - Diagnosis

MMHCC Cancer Models Database - Microsoft Internet Explorer

NATIONAL CANCER INSTITUTE

MMHCC the Mouse

Click folders to expand and select appropriate diagnosis or enter a search unable to find a match, enter it in the Other field below and click Submit.

Search:  Search

Match Case  Match Whole Words

- [-] Diseases of the Mouse Intestinal Tract (5)
  - [-] Neoplasms of the Mouse Intestinal Tract (3)
    - [+] Malignant Neoplasms of the Mouse Intestinal Tract (8)
    - [+] Benign Neoplasms of the Mouse Intestinal Tract (6)
      - [Gastro-Intestinal Intraepithelial Neoplasia](#)
    - [+] Preneoplastic Lesions of the Mouse Intestinal Tract (1)
    - [+] Benign Conditions of the Mouse Intestinal Tract (8)
    - [+] Hyperplasia of the Mouse Intestinal Tract (3)
    - [+] Inflammatory Diseases of the Mouse Intestinal Tract (3)



# Search Results

- ◆ Image annotations may include a detailed description, species, organ, diagnosis, strain, and image dimensions
- ◆ Researchers can view, zoom, and pan an image



**Cancer Models Database - Submission - Microsoft Internet Explorer**

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites History Mail Print Edit Discuss Real.com Messenger

Address <http://cancerimages.nci.nih.gov/calIMAGE/simpleresults.jsp?q=0>

**NATIONAL CANCER INSTITUTE** **calIMAGE**  
Cancer Image Server

Home Help New Search

**Search Criteria: = all**

Page 1 Show 1 - 5 of 244 | [First](#) | [Next](#) | [Last](#) |

	<p><b>Description:</b> ML6(Axras KO (+/-)) This slide is from a genetically engineered mouse with a ras knock out (Axras KO (+/-)). The lung contains multiple tumors. Note that these tumors are relatively darkly stained, have lobular, crisply delineated borders. The largest tumor (<a href="#">x=24455 y=7578 1:5</a>) is connected to the bronchus which has dysplastic cells (<a href="#">x=24189 y=6774 1:1</a>). A papillary part of the tumor extends into the lumen. The cells in this tumor tend to have large pleomorphic, hyperchromatic nuclei with a high mitotic rate and scanty, dark cytoplasm (<a href="#">x=23360 y=8110 1:1</a>). The level of differentiation varies from well defined glandular patterns (<a href="#">x=12553 y=26890 1:3</a>), papillary (<a href="#">x=14457 y=3840 1:3</a>) to a solid invasive pattern (<a href="#">x=17985 y=12200 1:2</a>). Diagnosis: Multiple Adenocarcinomas, lung.</p>
ML6.sid	<p><b>Species:</b> Mouse <b>Donator Reference:</b> Robert Cardiff <b>Gender:</b> Unknown <b>Image Width:</b> 45140 <b>Image Height:</b> 39984 <b>Unit Type:</b> Pixels <b>Magnification Level:</b> 7</p>
	<p><b>Description:</b> C1(09SMA292) This slide is from a GFAP-Cre;RbLoxP/LoxP;p53LoxP/LoxP mouse showing a massive brain tumor (<a href="#">x=12210 y=9015 1:16</a>) that is arising from the cerebellum (<a href="#">x=3919 y=11950 1:4</a>). The tumor appears in direct continuity with the granular layer (<a href="#">x=8042 y=13756 1:3</a>). The tumor has a population of poorly differentiated, relatively large cells with pleomorphic nuclei (<a href="#">x=11112 y=11977 1:1</a>) and is highly vascular (<a href="#">x=13370 y=8029 1:2</a>). Diagnosis: Large cell/anaplastic medulloblastoma.</p>
C1.sid	<p><b>Species:</b> Mouse <b>Donator Reference:</b> Robert Cardiff <b>Gender:</b> Unknown <b>Image Width:</b> 24608 <b>Image Height:</b> 17859 <b>Unit Type:</b> Pixels <b>Magnification Level:</b> 6</p>
	<p><b>Description:</b> C3(206A15) This slide is from a PDGF-B-Chain retrovirus-induced mouse showing a massive brain tumor (<a href="#">x=9903 y=4697 1:15</a>). The tumor has a population of poorly differentiated cells with vascular proliferation (<a href="#">x=13162 y=2679 1:4</a>) and palisading (<a href="#">x=12258 y=5272 1:4</a>). The tumor cells have hyperchromatic small nuclei with scattered tumor giant cells. Diagnosis: Glioblastoma multiforme, mouse brain.</p>
C3.sid	<p><b>Species:</b> Mouse <b>Donator Reference:</b> Robert Cardiff <b>Gender:</b> Unknown <b>Image Width:</b> 45361 <b>Image Height:</b> 63335 <b>Unit Type:</b> Pixels <b>Magnification Level:</b> 7</p>

Done Internet



# Organ Model Site Images

◆ Annotated images are available for specific organ models on the MMHCC emice web site

Emice website - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: [http://emice.nci.nih.gov/emice/mouse\\_models/organ\\_models/lung\\_models/murinecancer/murinemodels](http://emice.nci.nih.gov/emice/mouse_models/organ_models/lung_models/murinecancer/murinemodels) What's Related

BlackBerry Starpower: The Instant Message Secure Web Shop My Presario Compaq At Home Compaq Support Smart Update

[Classification and Staging](#)

- [Molecular Alterations](#)
- [Novel Therapeutics](#)

**Murine Lung Cancer**

- [Murine Models of Lung Cancer](#)
- [Classification of Murine Lung Tumors](#)

[References](#)

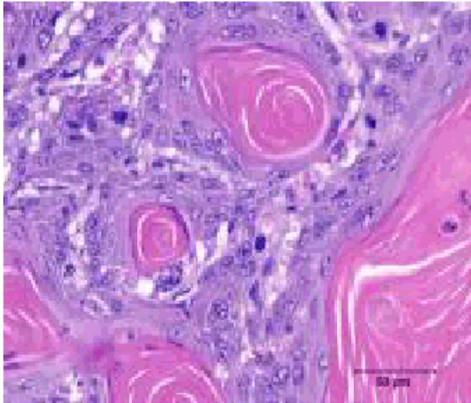
[Download report \(PDF\)](#)

**Protocols and Resources**

- [Lung Cancer Models](#)
- [Mouse Lung Images](#)
- [Human Lung Images](#)
- [Human Lung Pathways](#)
- [Human Lung Genes \(CGAP\)](#)
- [Mouse Lung](#)

sensitivity to lung cancer. Analysis of progeny from crosses between recombinant inbred (RI) strains derived from the sensitive A/J strain and the resistant C57BL/6J strain, suggested the existence of three pulmonary adenoma susceptibility (Pas) loci (55). Pas-1 was later identified by analysis of F2 progeny from a cross between strain A/J and the C3H/He resistant strain and was mapped to the distal region of chromosome 6 (24). Linkage analysis has demonstrated K-ras to be tightly linked to the Pas-1 locus, suggesting K-ras as a candidate for Pas-1 (47). Additional Pas loci have been mapped to chromosomes 9, 17 and 19 (13, 18). Furthermore, several numerous susceptibility to lung cancer (Sluc) loci have been identified by using a multilocus mapping method to analyze F2 mice generated from recombinant congenic strains (RCS). The Sluc loci are involved in complex genetic interactions that control susceptibility to the development of lung cancer (19, 20).

A wide variety of chemical carcinogens can induce pulmonary adenoma and adenocarcinoma formation in mice although they vary in their potencies (for a review on spontaneous and chemically induced mouse lung tumors see (53, 84, 86)). Some well characterized tumorigenic agents include urethane, metals, aflatoxin, tobacco smoke and tobacco smoke constituents including polyaromatic hydrocarbons and nitrosamines. Of note, the only murine model of squamous cell carcinoma existing to date is a carcinogen induced model resulting from the topical administration of NTCU twice a week for 35-40 weeks (71). The study of chemically induced lung tumors has provided insights into the histogenesis of murine lung tumors suggesting that murine pulmonary adenocarcinomas are derived from cells of the alveolar epithelium of the type II cell lineage, or from the bronchiolar epithelium of the Clara cell lineage (34, 70, 72, 92).



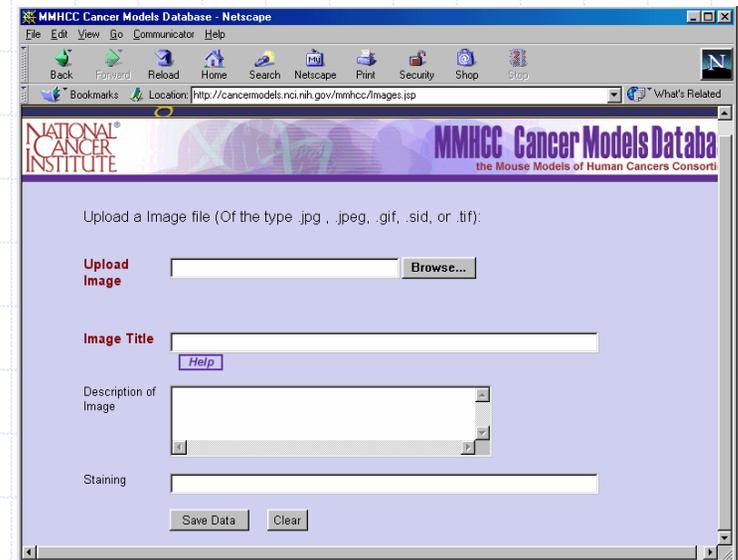
**Diagnosis:** Keratinizing squamous carcinoma of the lung.  
**Species:** Mouse  
**Boston reference set number:** LW071A

Document: Done

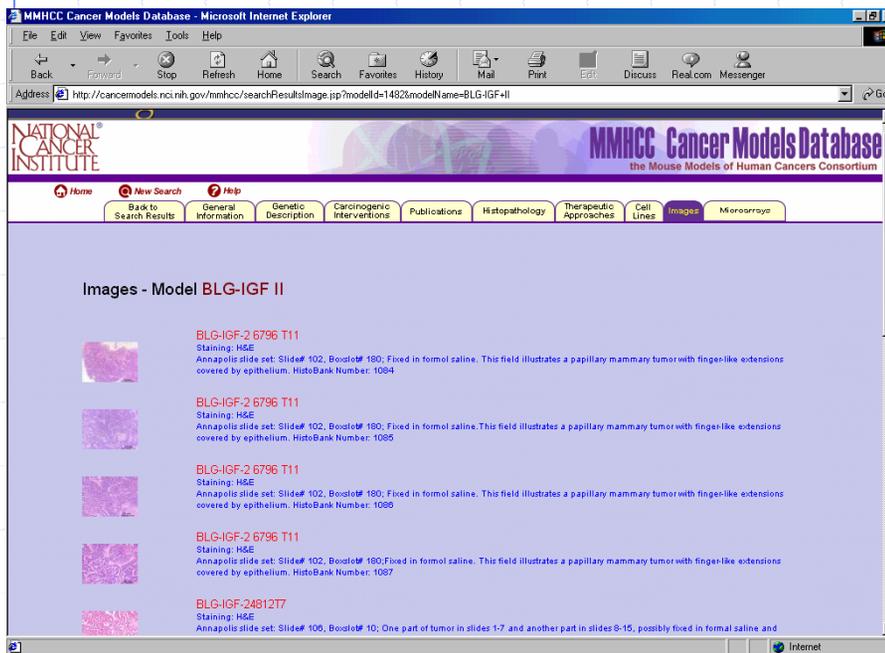


# Cancer Model Images

- ◆ Upload image constructs affiliated with their model via the Cancer Models Data Portal

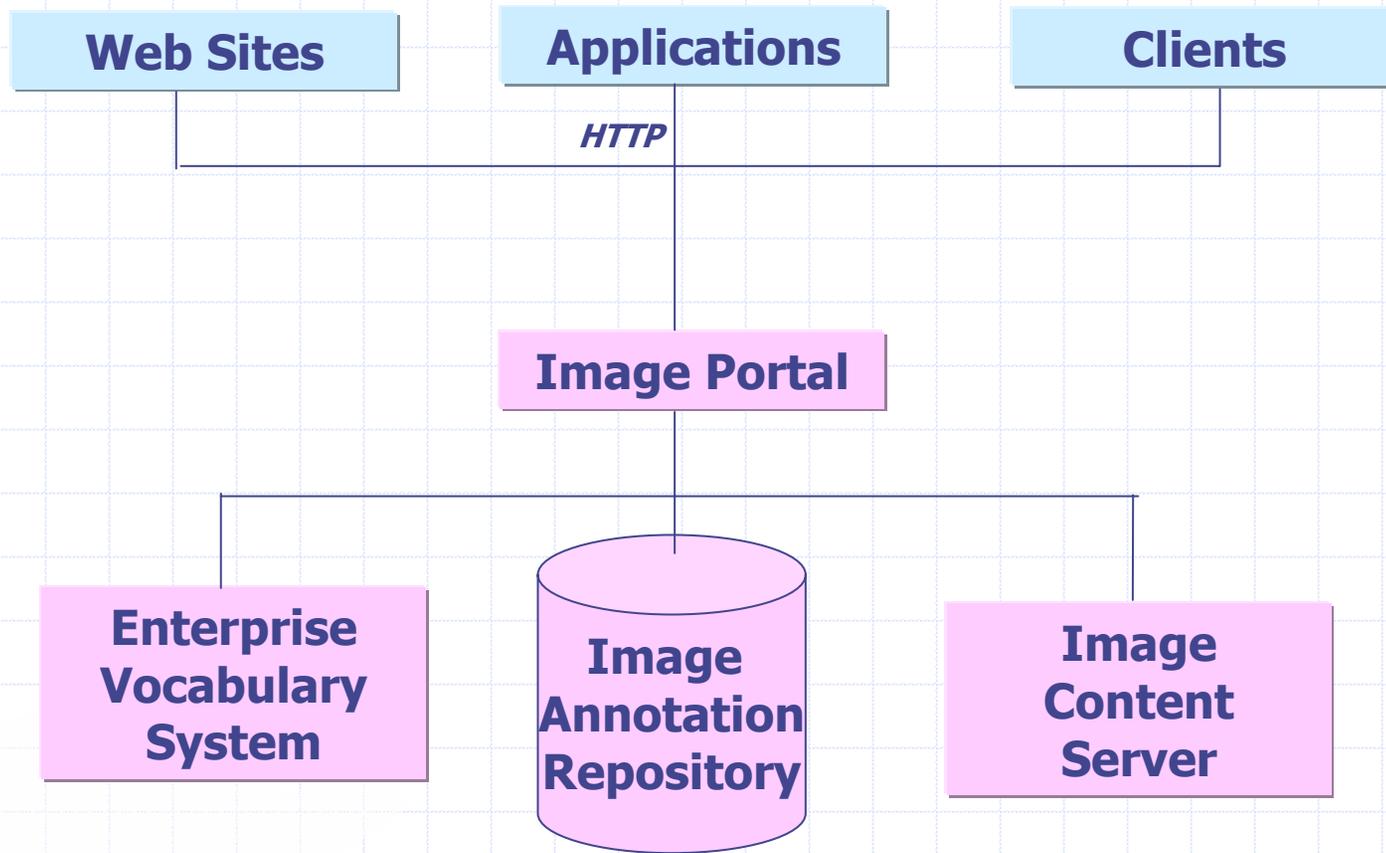


- Retrieve images associated with a model



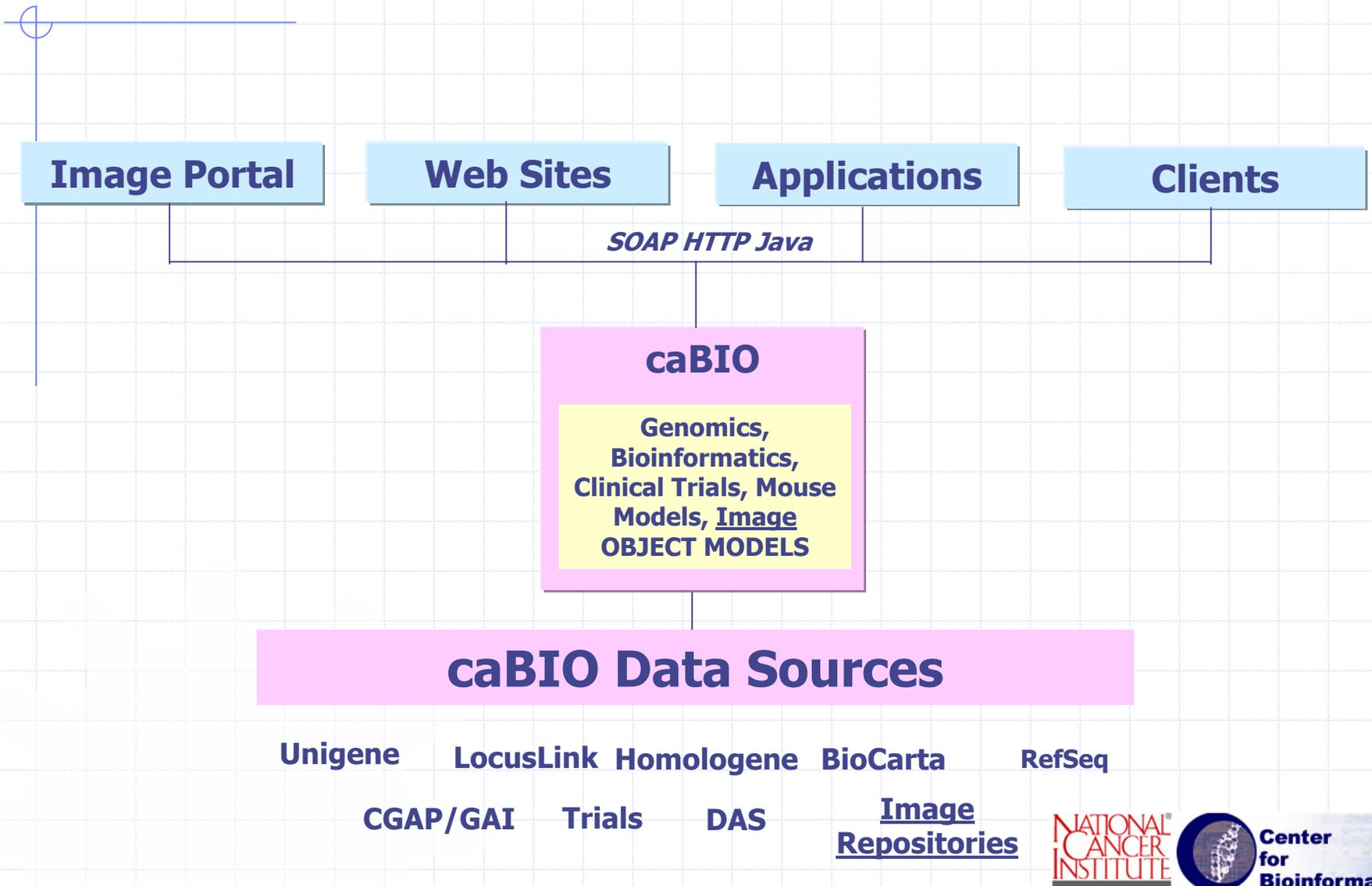


# Existing Architecture





# Future Architecture "Image Web Services"







# Image Annotation Standards

◆ To facilitate the sharing of histology images, a “minimal” set of image annotations are necessary

◆ Should leverage existing standards and may be derived from use cases for image retrieval and analysis

◆ Annotations should include parent-child relationships

```
C:\NCI\MMHCC\image.xml - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History Mail Print Edit Discuss
Address C:\NCI\MMHCC\image.xml
Links Agreement Customize Links Free Hotmail Headlines Home Instant Internet Instant Search My Presario News RealPlayer

<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE ImageAnnotation (View Source for full doctype...)>
- <ImageAnnotation>
- <Source>
  <Donator>Dr. John Doe</Donator>
  <Institution>University X</Institution>
</Source>
- <File>
  <Name>mp30</Name>
  <Location>http://www.universityx.com/catalogview.jsp?cat=mouse_p_prostate&=mp30.sid&reslvl=7.1
  &rgn=0,0,1,1</Location>
</File>
- <Dimensions>
  <Width>300</Width>
  <Height>300</Height>
  <Type>JPEG</Type>
  <Magnification>7</Magnification>
</Dimensions>
<Description>MP30 (01-0948SF) This slide illustrates the distribution of cytokeratin 8 (CK*) in lesions from in TRAMP male
  mouse using the immunoperoxidase technique to identify the antigen.</Description>
<Diagnosis>Intestinal metaplasia (probasin:rasxmx1), Mouse Prostate</Diagnosis>
<Publication ID="504">PubMed</Publication>
<Organ>prostate</Organ>
<Species>mouse</Species>
<Strain>Iodine</Strain>
<Gender>Male</Gender>
<Promoter>Probasin</Promoter>
- <Gene>
  <GeneSymbol>p53</GeneSymbol>
  <GeneSymbol>PTEN</GeneSymbol>
</Gene>
<Stain>H&E</Stain>
</ImageAnnotation>
```



# Imaging Technologies

- ◆ The NCICB is investigating imaging technologies to facilitate efficient image retrieval and annotation integration
- ◆ Imaging technologies include JPEG 2000, DICOM 3, and Image Content Servers
  - JPEG 2000 is a standard that defines a set of lossless (bit-preserving) and lossy compression methods for coding continuous-tone, bi-level, gray-scale, or color digital still images
  - DICOM 3 (Digital Imaging and Communications in Medicine) is the industry standard for the transfer of radiology images and other medical information between computers
  - Image Content Servers provide a mechanism to speed image transmission and improve image quality



# Image Contributors



- ◆ Dr. Robert Cardiff, UC Davis – Mouse Prostate Cancer Images
- ◆ Dr. William Weiss, Dr. Ken Aldape, UCSF – Nervous System Cancer Images
- ◆ Dr. Tyler Jacks, MIT – Lung Cancer Images
- ◆ Dr. Alex Nikitin, Cornell – Lung Cancer Images
- ◆ Dr. Norman Greenberg, Baylor – Prostate Cancer Images
- ◆ Dr. Scott Shappell, Vanderbilt – Prostate Cancer Images
- ◆ Dr. Marcus Bosenberg, Harvard – Skin Cancer Images