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*

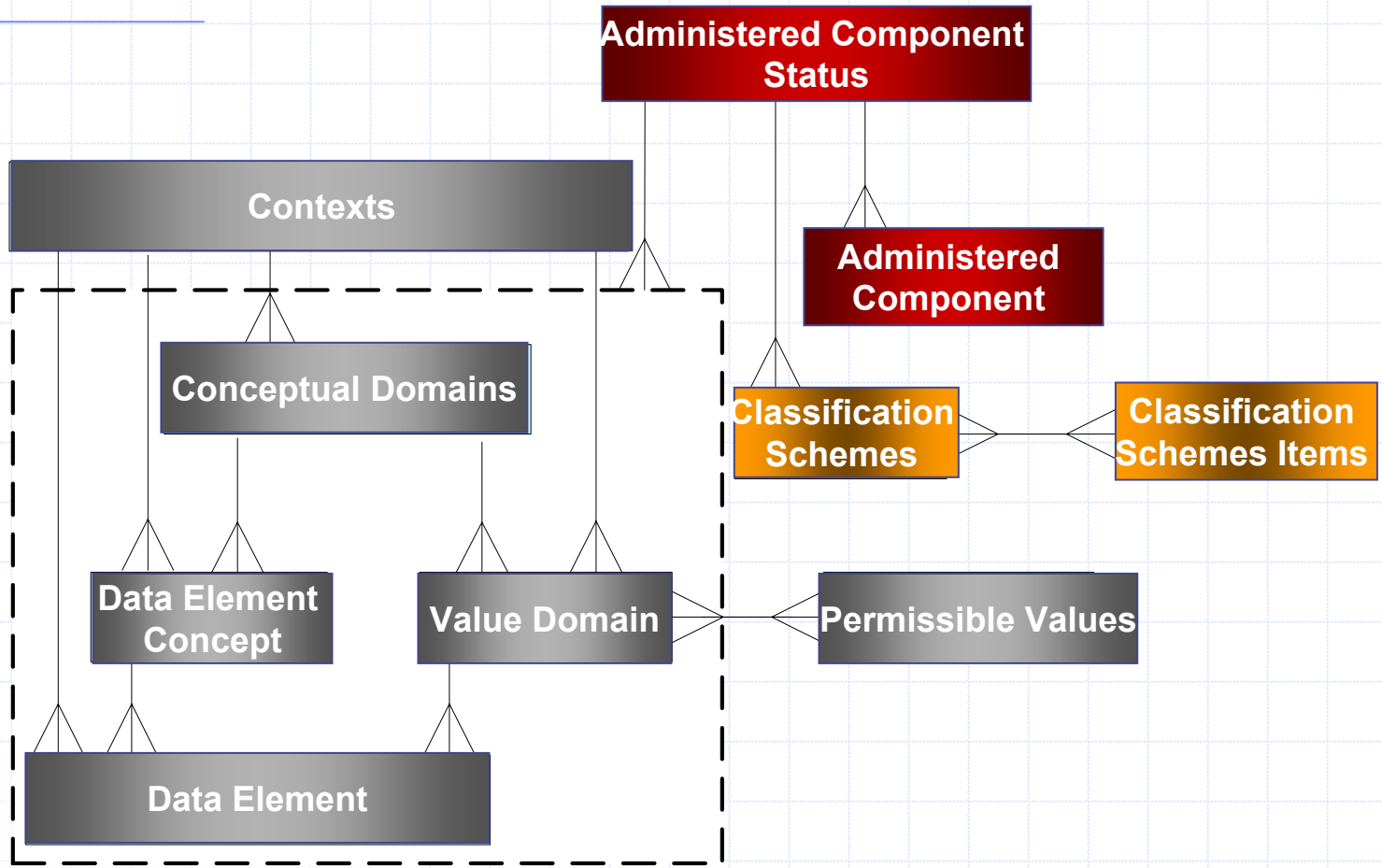
caCORE

- 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

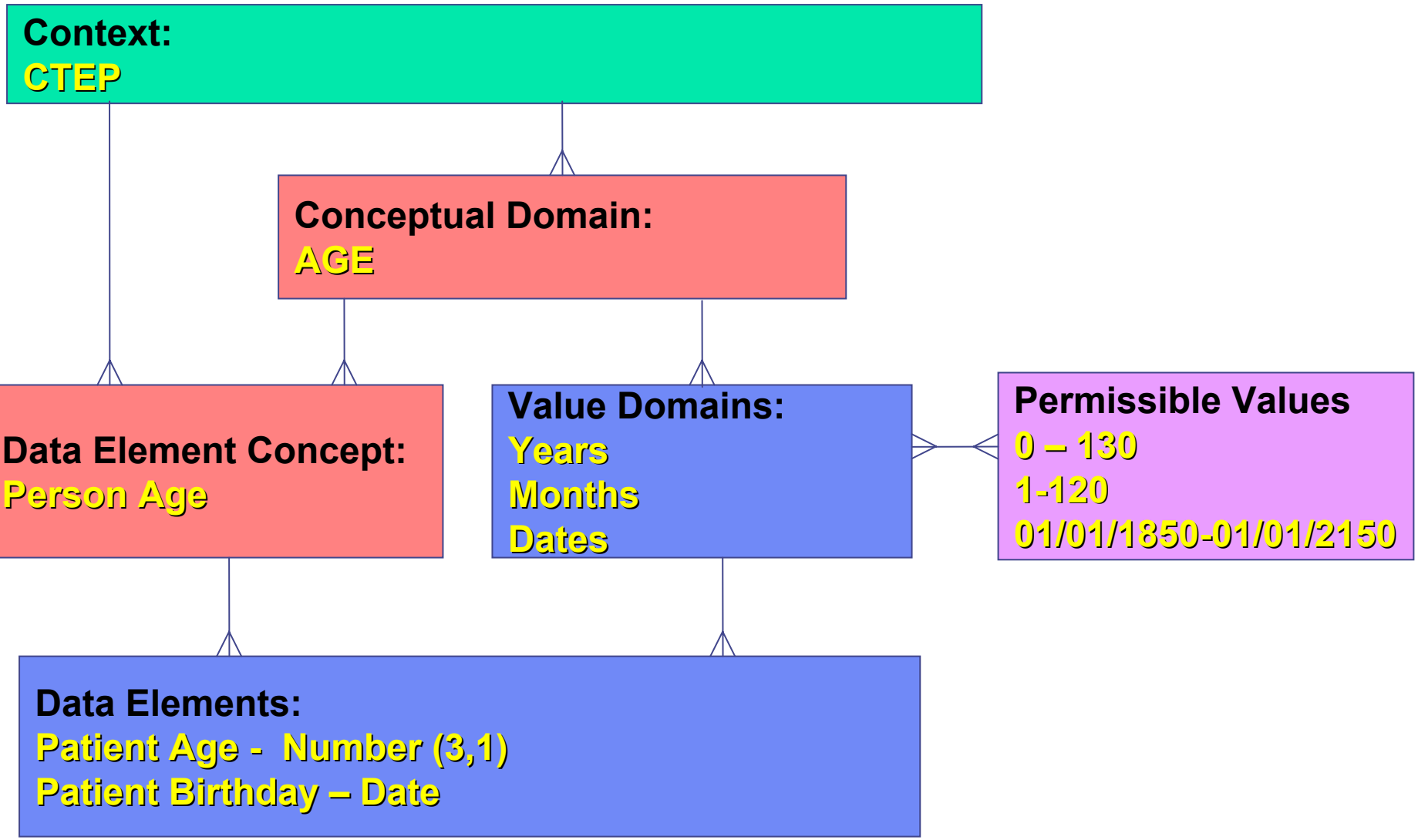
Cancer Data Standards Repository (caDSR)

- 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

CTEP

SPOREs
Specialized Programs of Research Excellence

NATIONAL CANCER INSTITUTE

Cancer Biomarkers Research Group

Early Detection Research Network

NATIONAL CANCER INSTITUTE

Center for Bioinformatics

NATIONAL CANCER INSTITUTE

Biomedical Imaging Program

Division of
CANCER PREVENTION

Different Curators for each Context

NATIONAL CANCER INSTITUTE

Center for Bioinformatics

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

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 Center for Bioinformatics

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

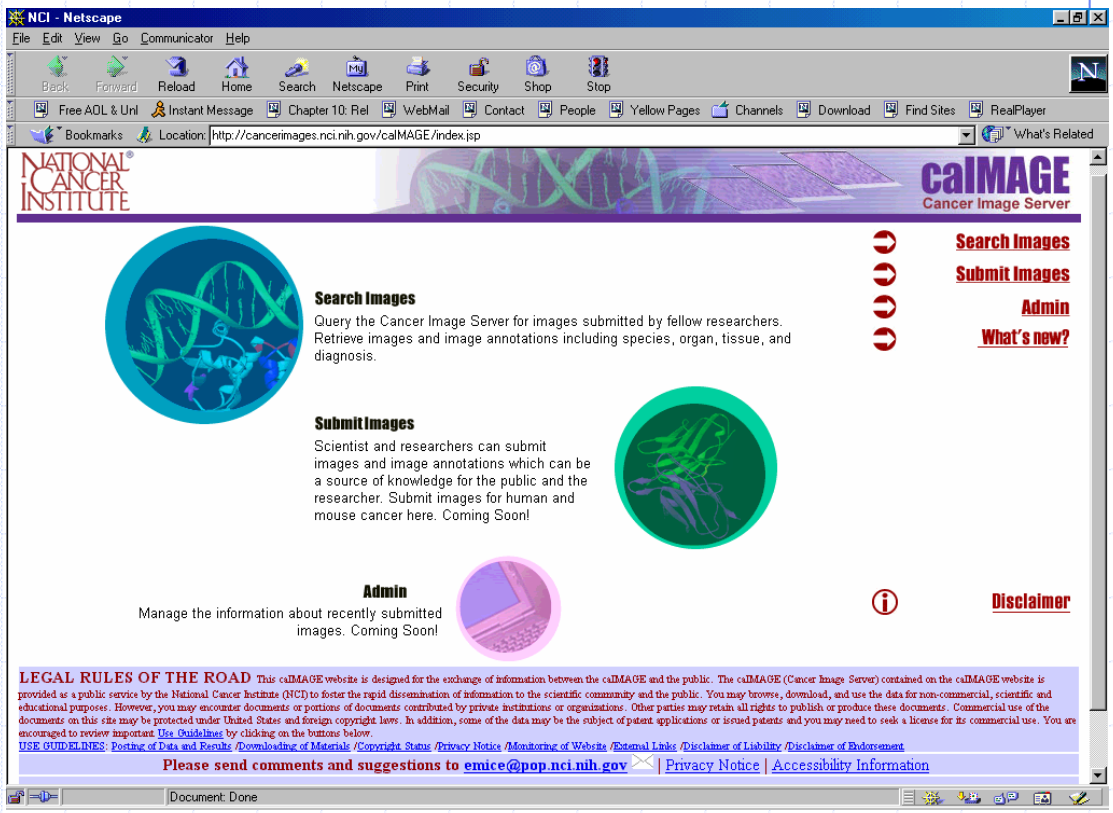
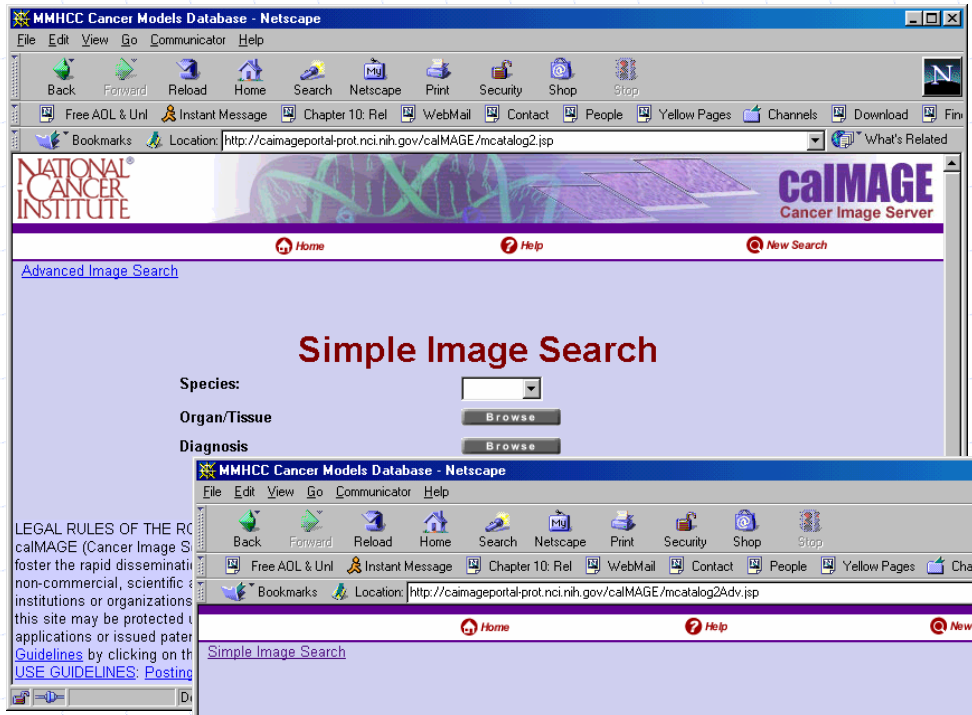


Image Search

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



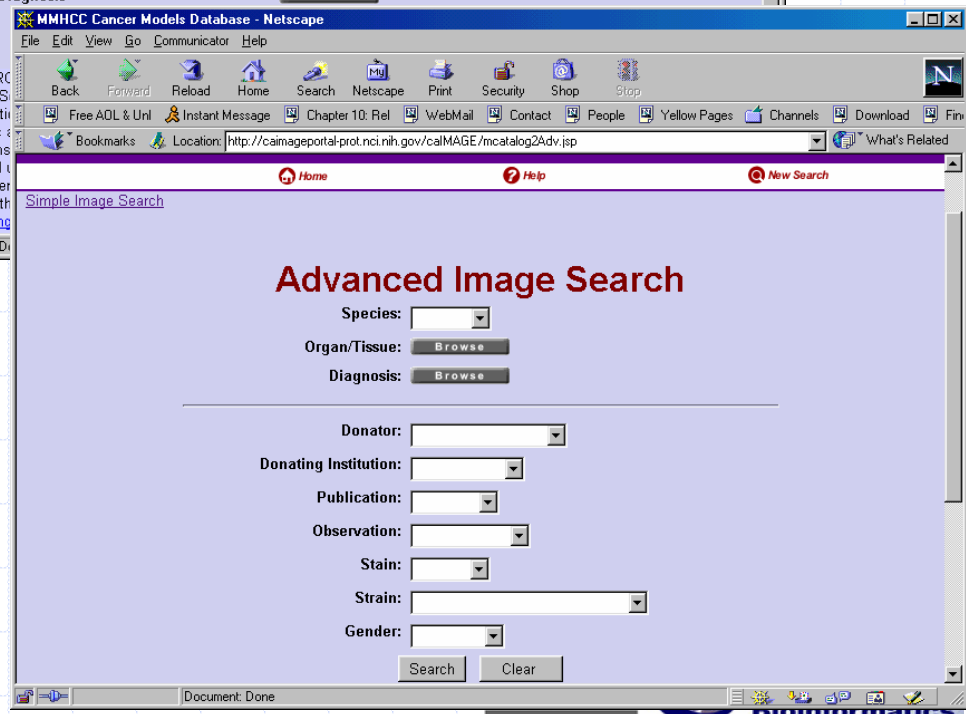
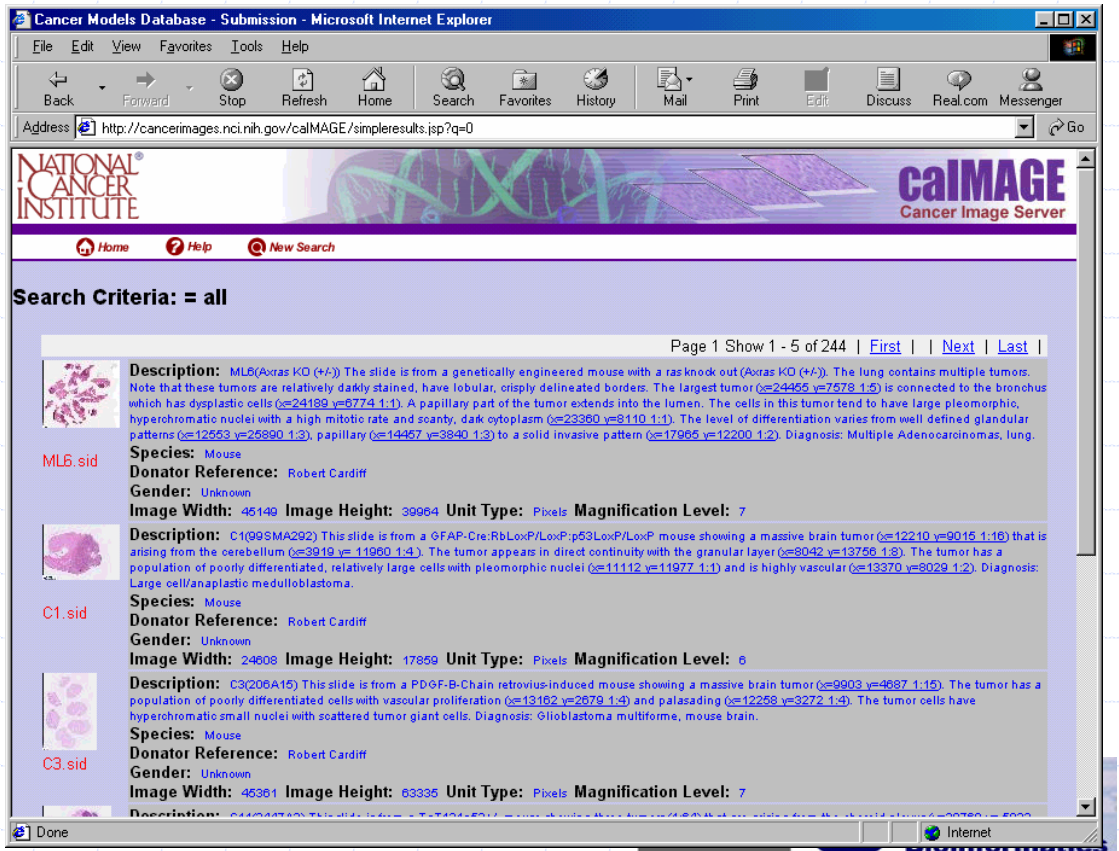


Image Vocabulary

- 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

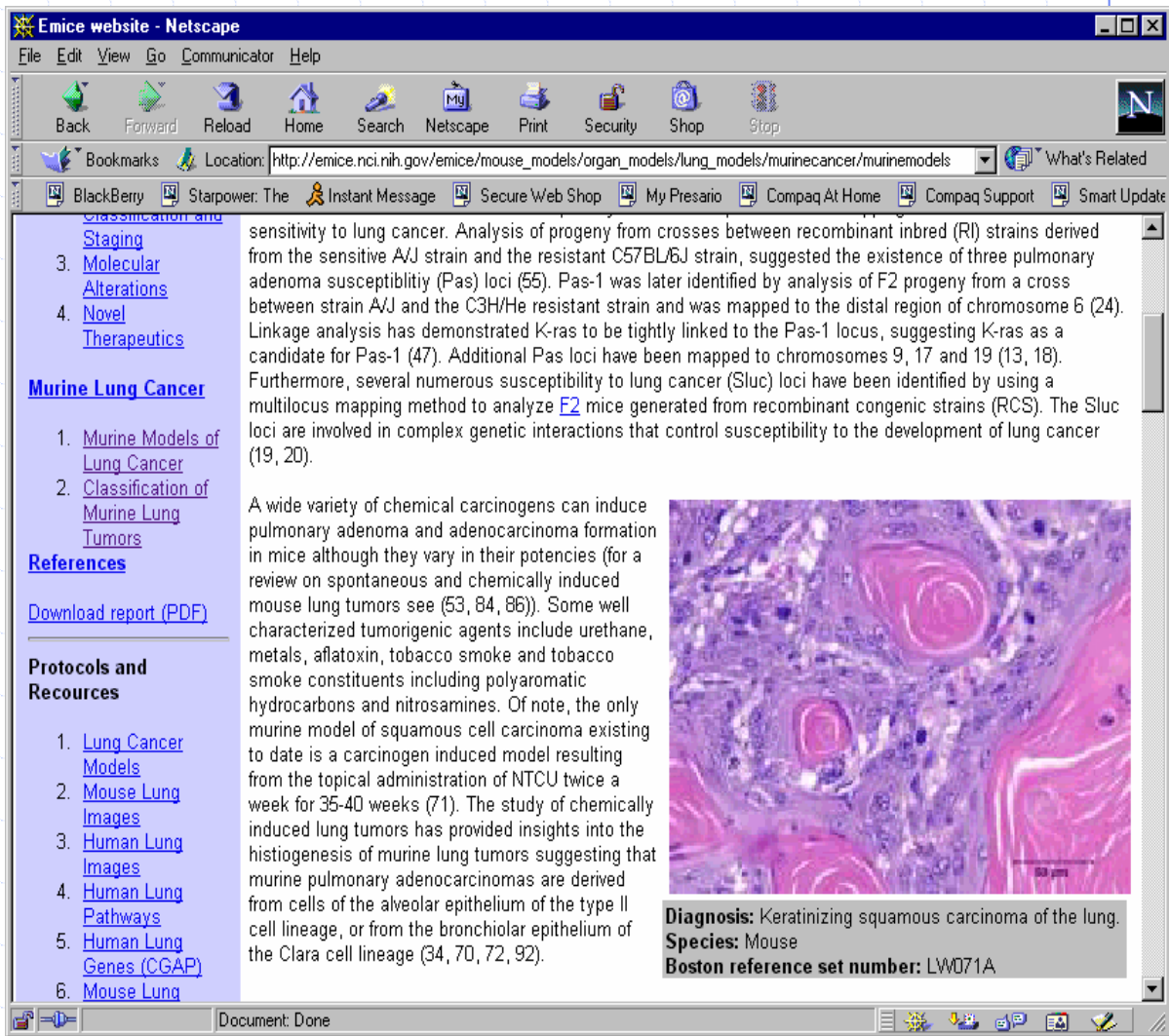
Search Results

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



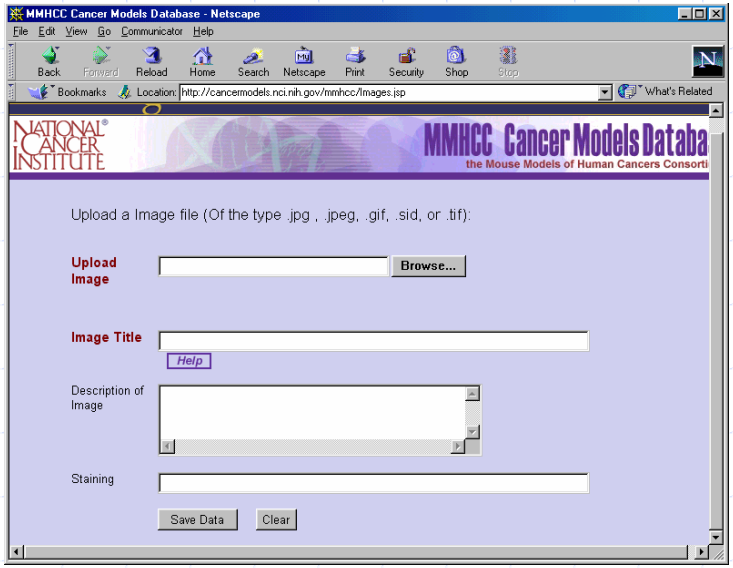
Organ Model Site Images

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



Cancer Model Images

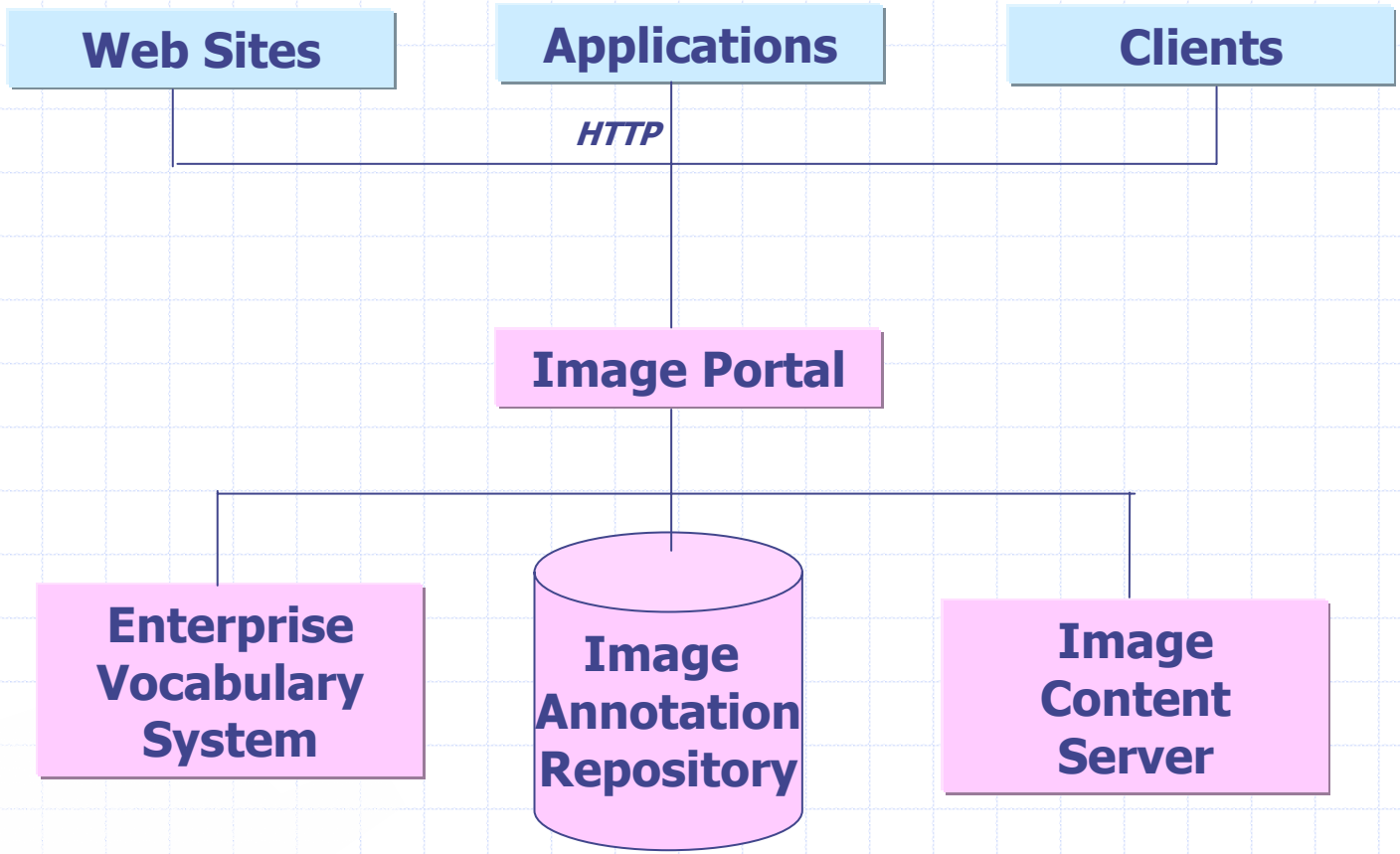
- Upload image constructs affiliated with their model via the Cancer Models Data Portal



- Retrieve images associated with a model

NATIONAL CANCER INSTITUTE Center for Bioinformatics

Existing Architecture



Future Architecture "Image Web Services"

Image Portal | Web Sites | Applications | Clients

SOAP HTTP Java

caBIO

Genomics, Bioinformatics, Clinical Trials, Mouse Models, Image OBJECT MODELS

caBIO Data Sources

Unigene LocusLink Homologene BioCarta RefSeq

CGAP/GAI Trials DAS Image Repositories

National Cancer Institute Center for Bioinformatics

Image Object Model

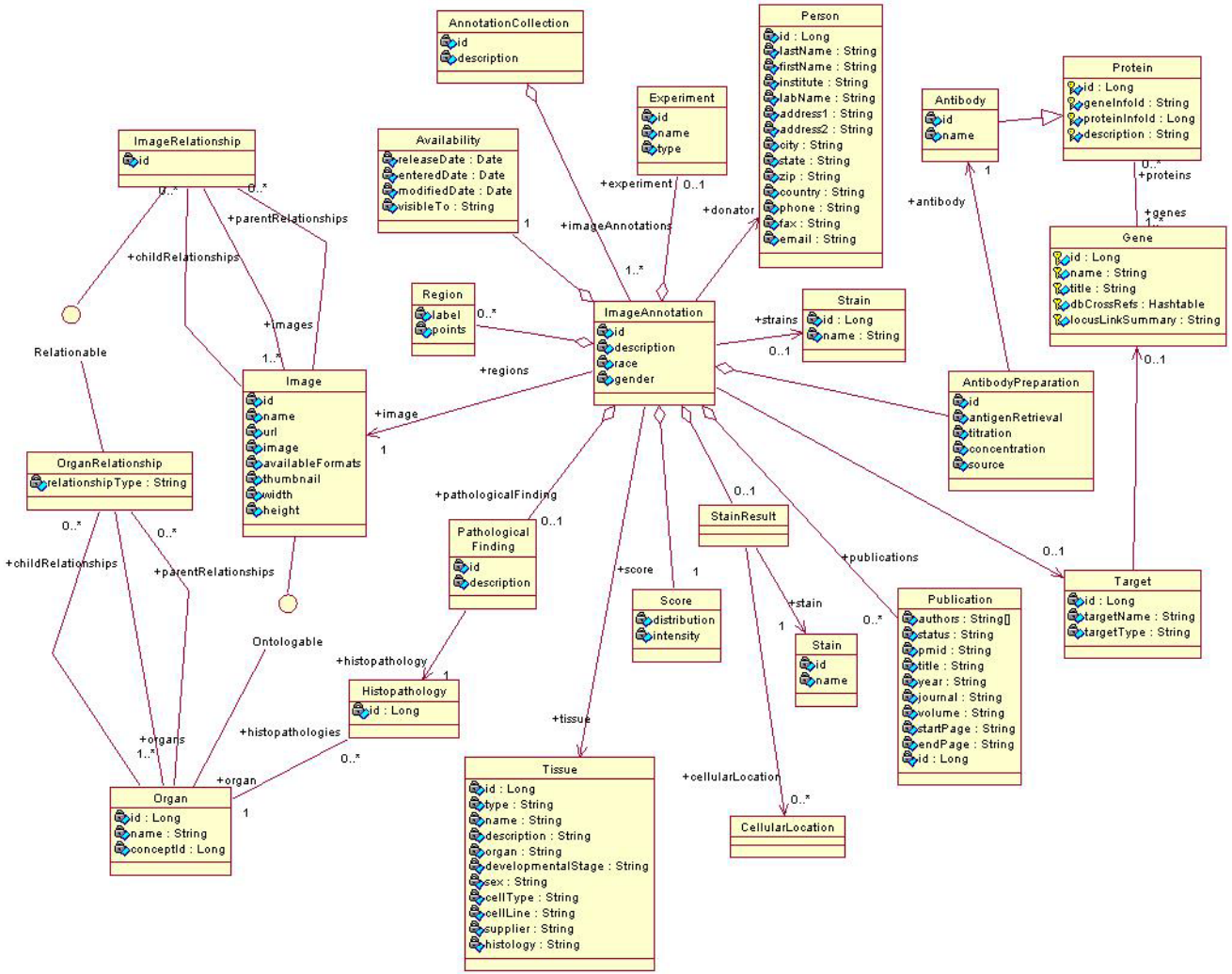
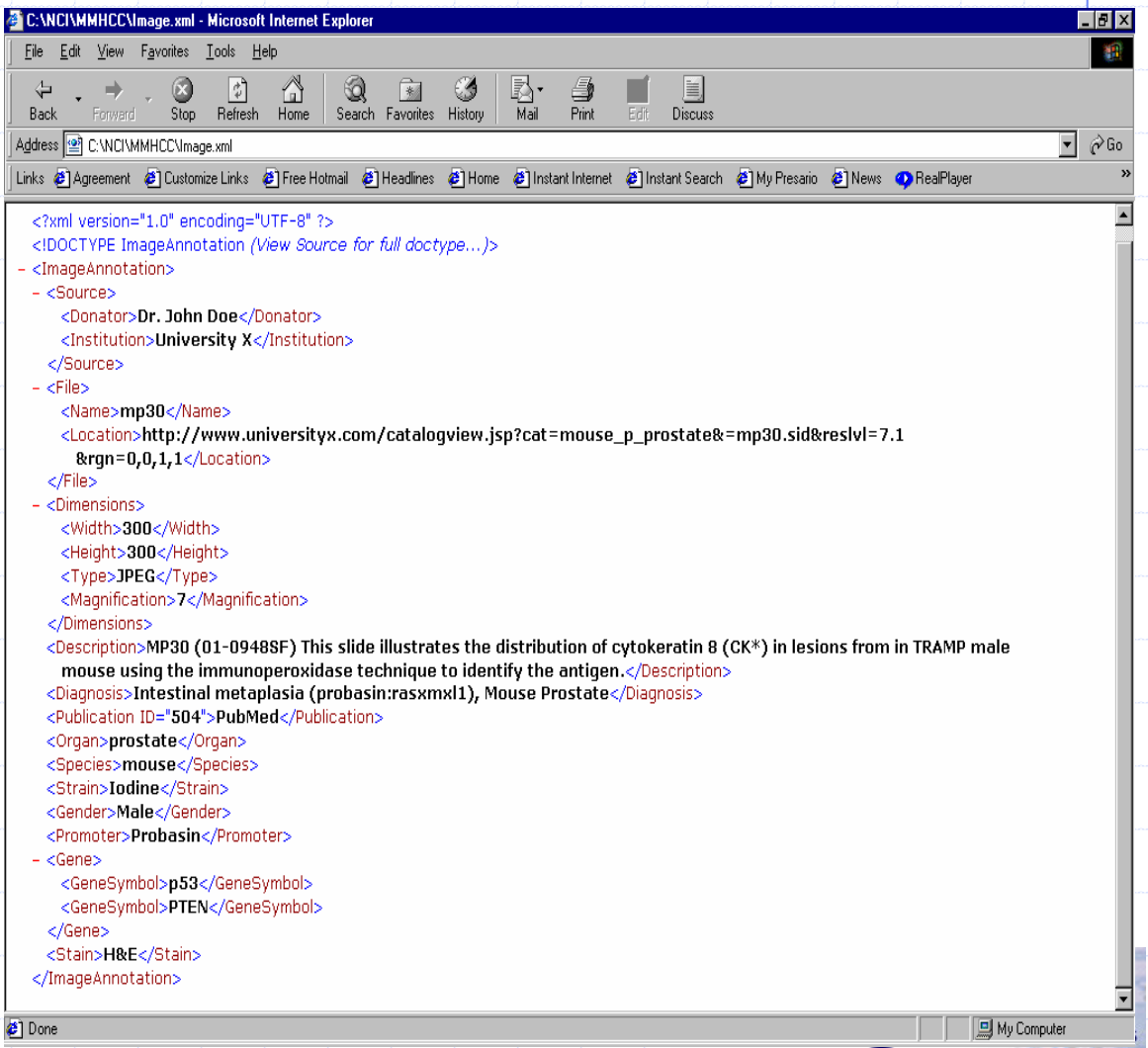


Image Annotation Standards

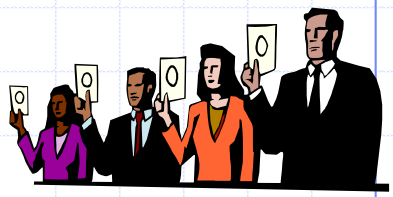
- To facilitate the sharing of histology images, a "minimal" set of image annotations are necessary
- Should leverage existing standards and may be derived from use cases for image retrieval and analysis
- Annotations should include parent-child relationships



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 transferal 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, Vanderbuilt – Prostate Cancer Images
- Dr. Marcus Bosenberg, Harvard – Skin Cancer Images