Agenda

- Overview of caMOD and its goals
- Salient Features
- Technical approach
- caBIG compatibility
- Q & A

Goals

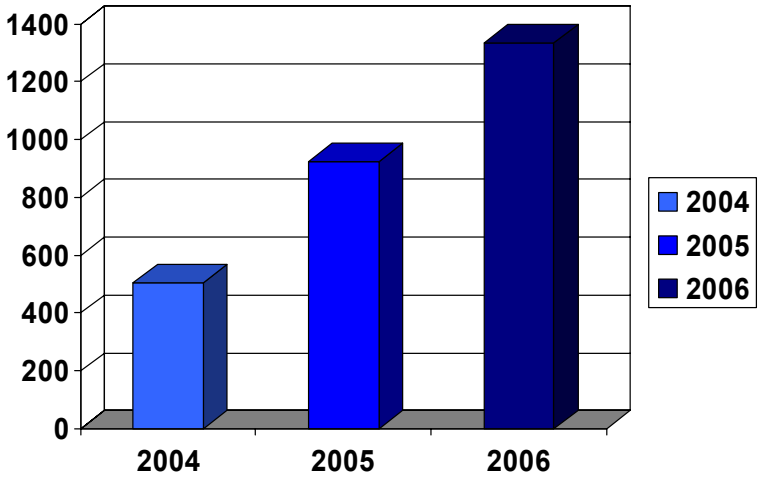
1. Cancer models that recapitulate many aspects of the genesis, progression, and clinical course of human cancers are valuable resources to cancer researchers engaged in a variety of basic, translational, clinical, and epidemiological investigations.

   To provide an application and infrastructure supporting cancer models of diverse species to help researchers understand the genesis, progression, and clinical course of human cancer.

2. The NCI Mouse Models of Human Cancers Consortium (MMHCC) is a collaborative program designed to derive and characterize animal models, and to generate resources, information, and innovative approaches to the application of animal models in cancer research.

   Provide a collaborative repository of internal and publicly available cancer models data and images supporting translational research.

3. Strive for a comprehensive knowledge source providing a wealth of information on cancer models, some pertinent aspects of which may never be available in a publication.



Avg # of Unique Visitors / Month

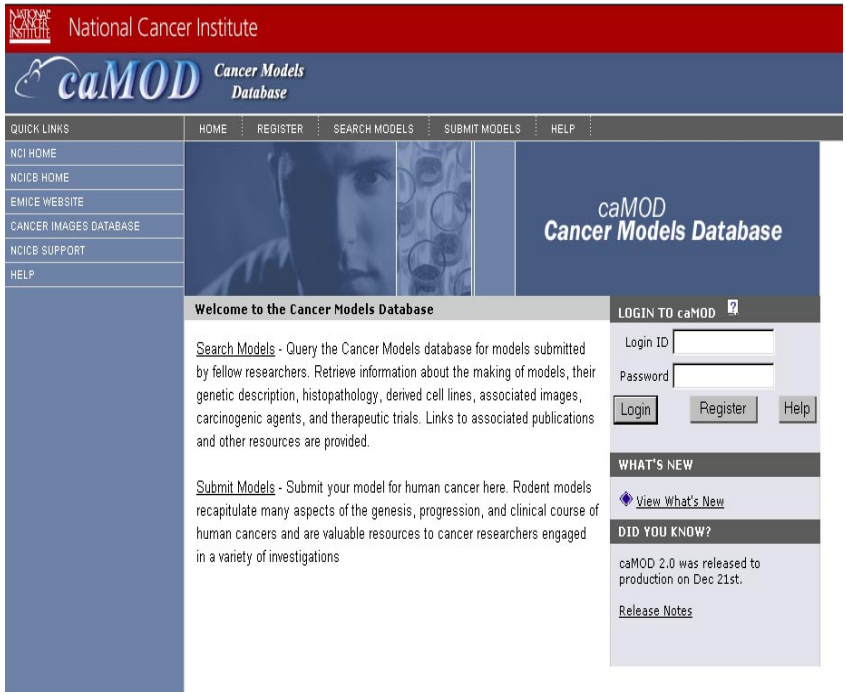
The cancer models database (caMOD) is a web-based resource that provides information about animal models for human cancer to the public research community

History

- January 2000
  - Prototype is presented during the Mouse Models of Human Cancers (MMHCC) Steering Committee Meeting
  - MMHCC adopts the Cancer Models Database (caMOD) as one of their initiatives
- July 2000
  - NCICB assumes responsibility for caMOD
- Spring 2001
  - caMOD 1.0 released (2-tier application)
- 2001 – 2005
  - Based on user feedback the application is constantly updated and improved.
- December 2005
  - caMOD 2.0 released (n-tier application, based on caBIG compliance guidelines)

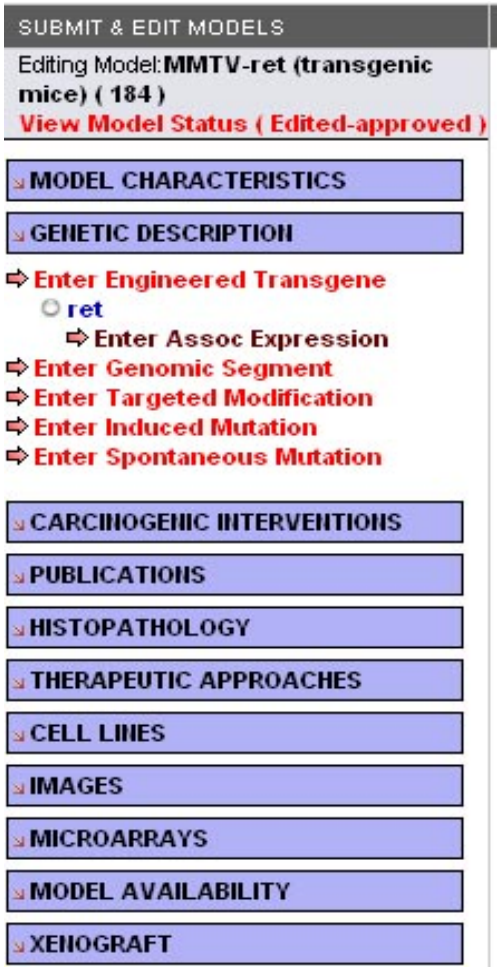
Cancer Models Database v2.0

- **Submission**--Data in caMOD are extracted from the public scientific literature by curators and verified by the scientists who generated or worked with the models, or they are directly submitted by scientists.
- **Search**--Users can retrieve information about the making of models, their genetic descriptions, histopathology, derived cell lines, associated images, carcinogenic interventions, microarray data, and therapeutic trials in which the models were used. caMOD provides links to PubMed for associated publications and other resources such as mouse repositories, detailed information about altered gene, pathway affected, and information about human clinical trials that utilize the same compounds as the pre-clinical trials in the animal models.
- **System Function Administration**--The Admin function provides services for user registration, review of submitted models and database management.



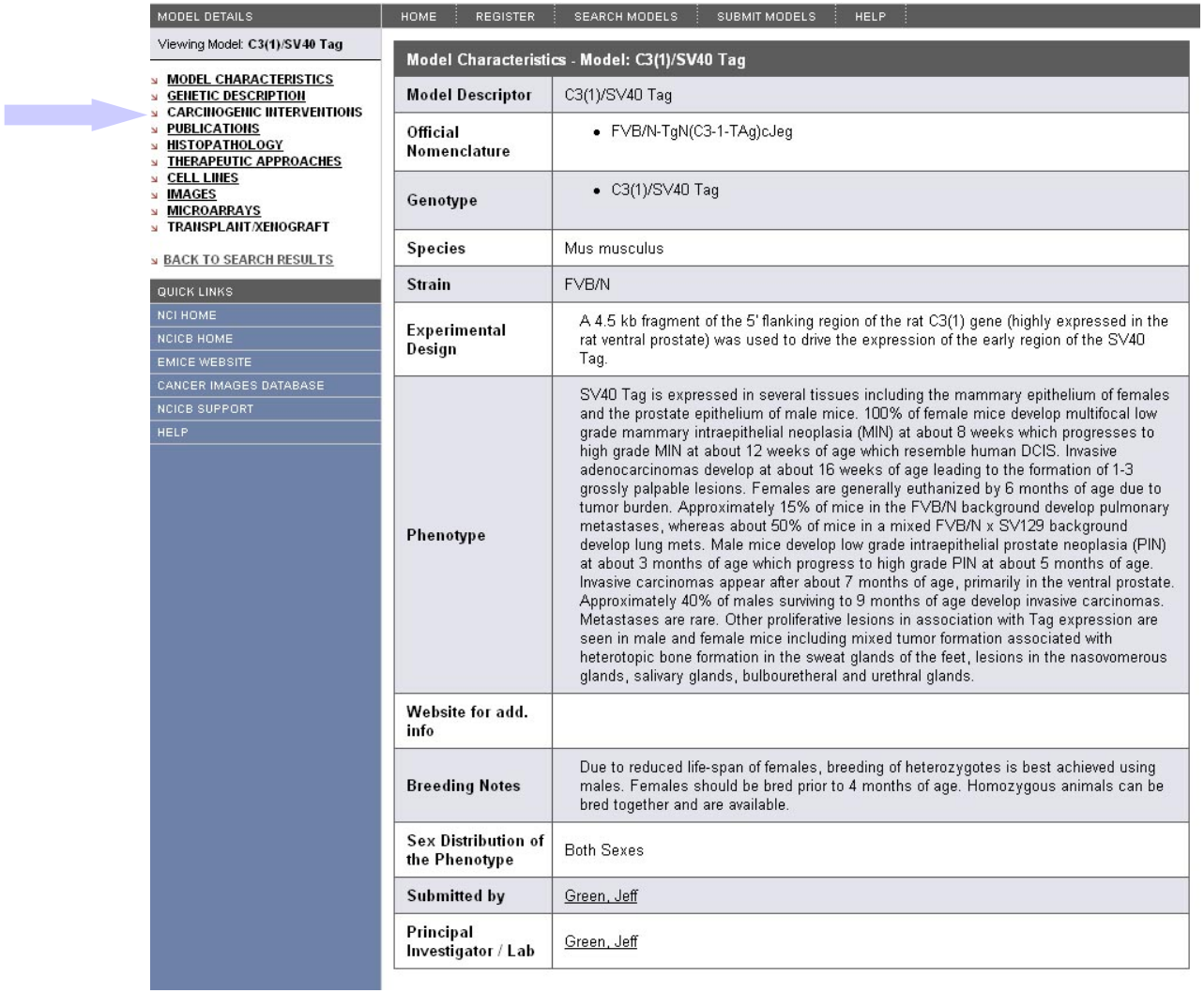
Navigating the submission pages

Most parts contain multiple pages



Multiple entries per category are possible

Search Result Detail Pages



caGRID Demo

- http://cagrid-browser.nci.nih.gov/

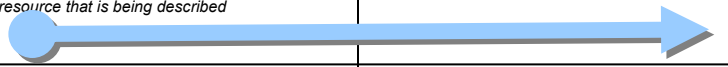
```
<caBIGXMLQuery name="my-query">
  <Target name="gov.nih.nci.camod.domain.AnimalModel">
    <Objects name="gov.nih.nci.camod.domain.AnimalModel">
      <Property name="id" value="3" />
    </Objects>
  </Target>
</caBIGXMLQuery>
```

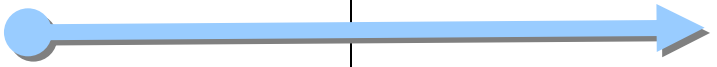
Interoperability

- caMOD has been designed, architected and constructed to facilitate interoperability with other systems, following caBIG guidelines.
- Information Providers to caMOD:
  - caBIO to retrieve gene info and clinical trials info through remote API
  - EVS to provide concept codes and preferred descriptions for concepts through caBIO EVS API
  - PubMED
  - Jackson Laboratory Resources
  - NCI's Developmental Therapeutics Program
  - caArray to store microarray data
  - caIMAGE server to store images
- Information Consumers: caMOD provides information to other systems
  - CMAP
  - BioGopher
  - Websites such as eMice references specific models in caMOD
  - caELMIR (future)

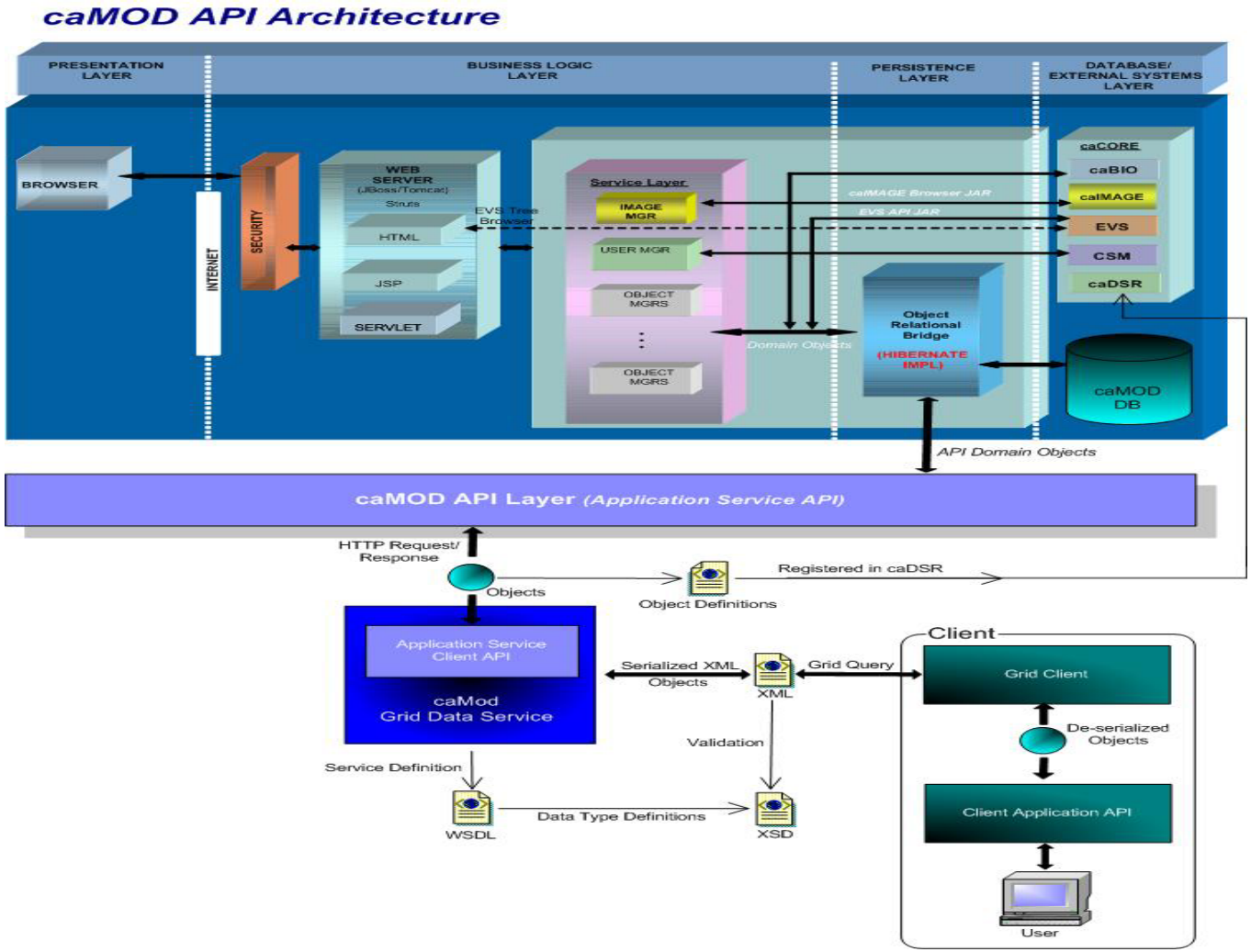
Evolution to caBIG Compliance

Maturity Model	Legacy	Bronze	Silver	Gold
Programming and Messaging Interfaces	*- No programmatic interfaces to the systemare available. Only local data files in a custom format can be read* *- Data transfer mechanisms implemented only on an ad hoc basis*	*- Programmatic access to data from an external resource is possible*	*- Well-described API's approved by the caBIG Architecture workspace provide access to data in the form of data objects that are instances of classes represented by a domain model* *- Electronic data formats reviewed and approved by the caBIG Architecture Workspace are supported for both input to and output from the system* *- Messaging protocols approved by the caBIG Architecture Workspace are supported wherever messaging is indicated by the use cases*	*- All features of Silver, plus:* *- Service-oriented components produce or use resources in the form of grid services that use XML as the primary interchange format.* *- Interoperable with caGrid data grid architecture being developed by caBIG Architecture Workspace- Other features to be determined by caBIG Architecture workspace*
Vocabularies / Terminologies & Ontologies	*- Free text used throughout for data collection*	*- Use of publicly accessible controlled vocabularies as well as local terminologies* *- Terminologies must include definitions of terms that meet caBIG VCDE workspace guidelines*	*- Terminologies reviewed and validated by the caBIG VCDE Workspace used for all appropriate data collection fields and attributes of data objects* *- Term definitions must meet VCDE Workspace guidelines*	*- All features of Silver, plus:* *- Full adoption of caBIG terminology standards as approved by the VCDE workspace. Terminologies must be available through a caGrid service*
Data Elements	*- No Structured metadata is recorded*	*- Data element descriptions are maintained with sufficient definitional depth to enable a subject matter expert to unambiguously interpret the contents of the resource without contacting the original investigator* *- Data elements are built using controlled terminology - Metadata is stored and publicized in an electronic format that is separate from the resource that is being described*	*- Common Data Elements (CDEs) built from controlled terminologies and according to practices validated by the VCDE workspace are used throughout* *- CDEs are registered as ISO/IEC 11179 metadata components in the caBIG Context of the cancer Data Standards Repository (caDSR)*	*- All features of Silver, plus:* *- CDEs designated as caBIG Standards by the VCDE workspace are used* *- Metadata is advertised and discoverable via the caGrid services registry*
Information Models	*- No model describing the system is available in electronic format*	*- Diagrammatic representation of the information model is available in electronic format*	*- Object-oriented domain information models are expressed in UML as class diagrams and as XMI files, and are reviewed and validated by the VCDE Workspace*	*- All features of Silver, plus:* *- Information models are harmonized across the caBIG Domain Workspaces*





Architecture



caMOD APIs

- A read-only non-remote API exists for the retrieval of cancer models called “eQBE” from the common-persistence package. Extending the foundation of the concept of basic QBE in Hibernate, eQBE provides a mechanism for object searching based on example object graphs using the objects’ property values as criteria for a query. Used internally in caMOD.
- Using the caCore SDK Toolkit, the external caMOD API Layer is built upon the generated framework and domain objects (API Domain Objects) along with the caCore Application Service remote interface infrastructure.
- Integration of the caMOD application into the caGrid infrastructure includes the deployment of the caMOD Grid Data Service node and registration to the caBIG Index Service.

Example of internal caMOD Java API

For the AnimalModel class defined in caMOD with eQBE we can query properties of member objects as follows.

```
Person thePI = new Person();
thePI.setFirstName("Jane");
thePI.setLastName("Doe");
AnimalModel theQBEModel = new AnimalModel();
theQBEModel.setPrincipalInvestigator(thePI);
List theMatchingModels = Search.query(theQBEModel);
```

The result List of animal models will include all AnimalModel objects whose associated Person object (through the PrincipalInvestigator link) has a firstName property value set to “Jane” and a lastName property set to “Doe”.

By default, Search.query will apply an equals evaluation to the property values.

Example of external caMOD remote Java API

For the AnimalModel class defined in caMOD we can query based on properties of the object as follows.

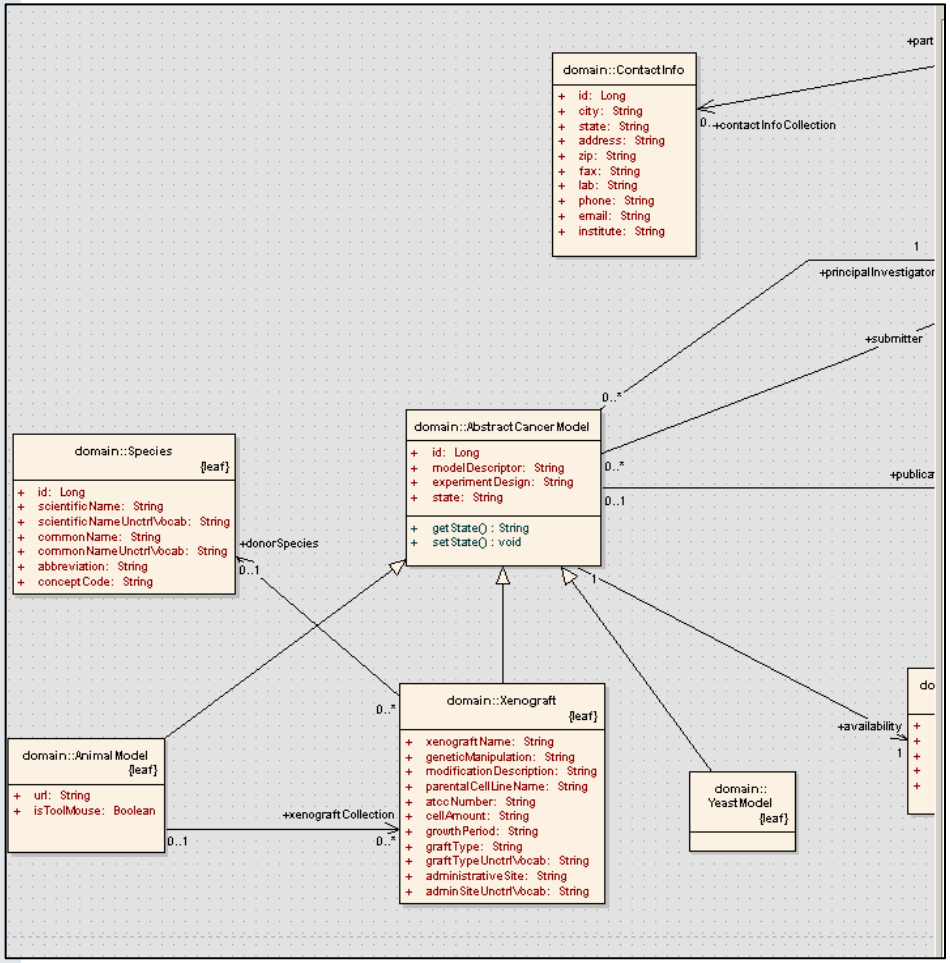
```
ApplicationService appService = ApplicationServiceProvider.getApplicationService();
 try {
     AnimalModel theAnimalModel = new AnimalModelImpl();
     theAnimalModel.setModelDescriptor("ARF+/-Emu-myc");

     List theAnimalModels = appService.search(AnimalModel.class, theAnimalModel);
     for (Iterator theIterator = theAnimalModels.iterator(); theAnimalModels.hasNext();)
     {
         AnimalModel theAnimalModel = (AnimalModel) theIterator.next();
         System.out.println("Id: " + theAnimalModel.getId() + "\n" + "\tType:" +
             theAnimalModel.getModelDescriptor() + "\n");
     }
  }
  catch (Exception e) {
      e.printStackTrace();
  }
```

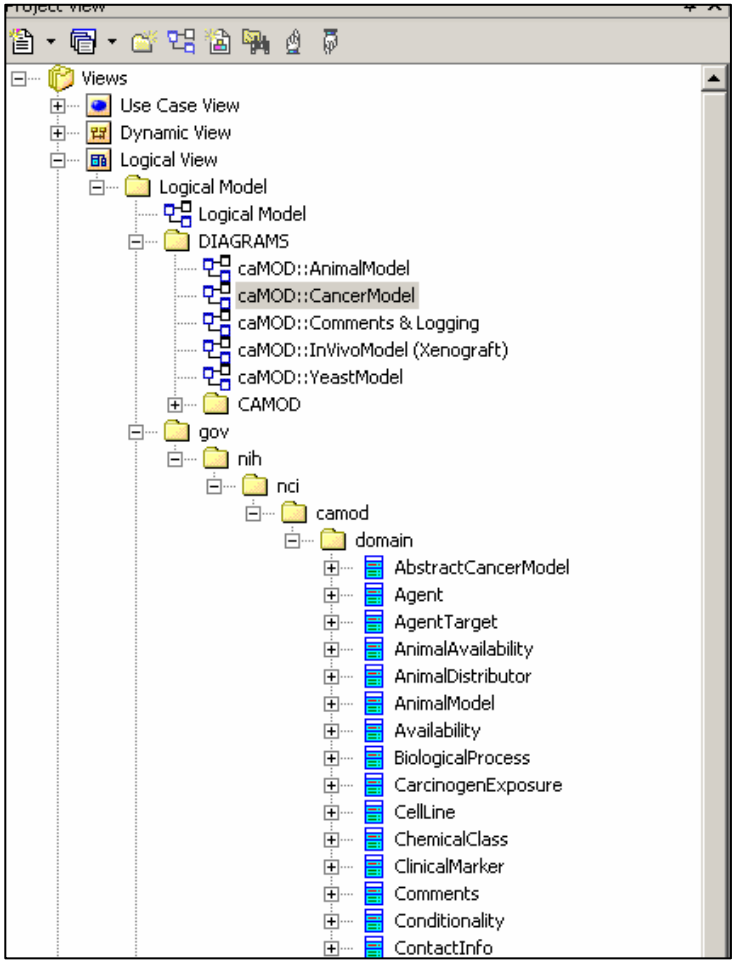
The result List of animal models will include all AnimalModel objects whose modelDescriptor property is equal to “ARF+/-Emi-mvc”.

caMOD Grid Node

- The caMOD Grid Data Service node is generated using the caGrid Toolkit.
- The Data Type Definitions extracted from the caDSR registered caMOD model are used to build an XML schema or XSD that is used to define the input and output data types of the Grid services.
- The XSD is registered in the Global Model Exchange (GME) and used to describe the contract of the grid service and to validate the XML serialization of the API Domain Objects during data requests.
- Applications communicate with the caMOD Grid Data Service node following the caGrid services protocol.
- The caMOD API domain objects are serialized/deserialized by the caMOD Grid Data Service to and from XML adhering to the XML Schema (XSD) that is registered in the Global Model Exchange (GME)

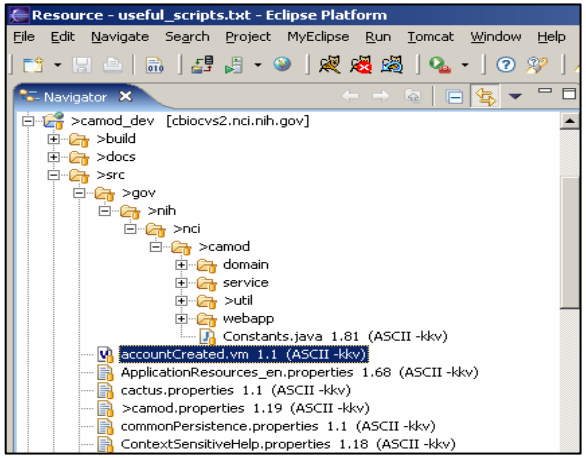
UML Model

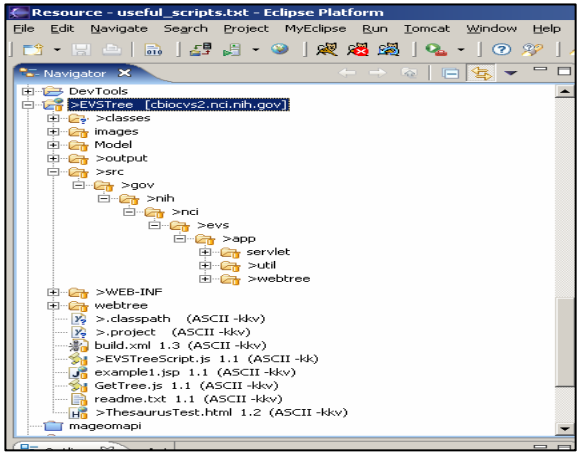

domain::ContactInfo
+ id: Long
+ city: String
+ state: String
+ address: String
+ zip: String
+ fax: String
+ lab: String
+ phone: String
+ email: String
+ institute: String
+part
0..*
+contactInfoCollection
1
+principalInvestigator
+submitter
0..*
domain::AbstractCancerModel
+ id: Long
+ modelDescriptor: String
+ experimentDesign: String
+ state: String
+ getState() : String
+ setState() : void
0..*
+publica
0..1
domain::Species
{leaf}
+ id: Long
+ scientificName: String
+ scientificNameUnctrlVocab: String
+ commonName: String
+ commonNameUnctrlVocab: String
+ abbreviation: String
+ conceptCode: String
+donorSpecies
0..1
1
0..*
domain::Xenograft
{leaf}
+ xenograftName: String
+ geneticManipulation: String
+ modificationDescription: String
+ parentalCellLineName: String
+ atccNumber: String
+ cellAmount: String
+ growthPeriod: String
+ graftType: String
+ graftTypeUnctrlVocab: String
+ administrativeSite: String
+ adminSiteUnctrlVocab: String
+availability
1
domain::Animal Model
{leaf}
+ url: String
+ isToolMouse: Boolean
+xenograftCollection
0..1
0..*
domain::
Yeast Model
{leaf}


Views
Use Case View
Dynamic View
Logical View
Logical Model
Logical Model
DIAGRAMS
caMOD::AnimalModel
caMOD::CancerModel
caMOD::Comments & Logging
caMOD::InVivoModel (Xenograft)
caMOD::YeastModel
CAMOD
gov
nih
nci
camod
domain
AbstractCancerModel
Agent
AgentTarget
AnimalAvailability
AnimalDistributor
AnimalModel
Availability
BiologicalProcess
CarcinogenExposure
CellLine
ChemicalClass
ClinicalMarker
Comments
Conditionality
ContactInfo

UML Model

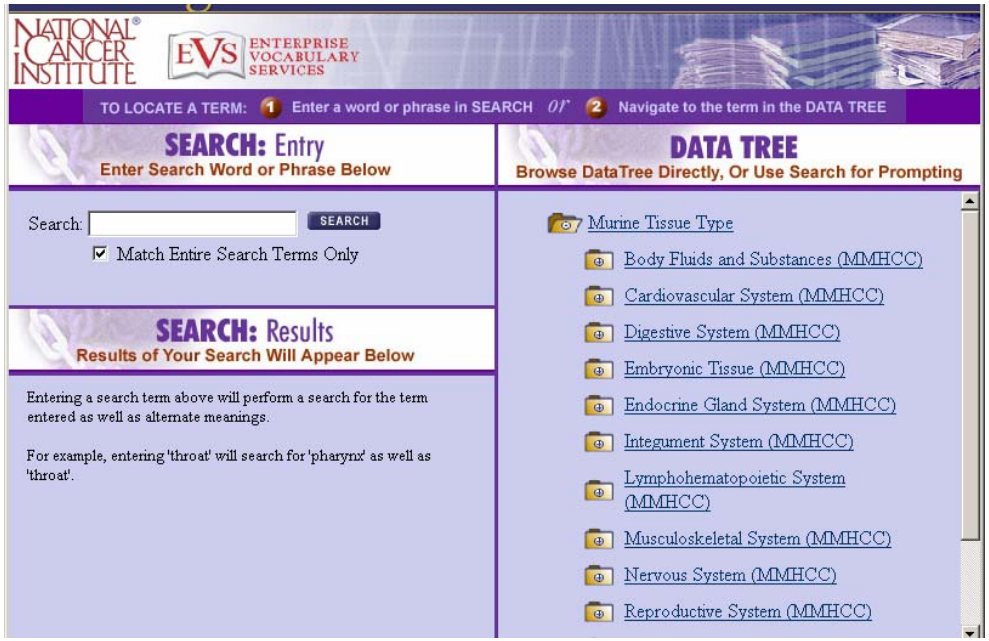
- 76 Domain objects categorized under
  - caMOD::AnimalModel
  - caMOD::CancerModel
  - caMOD::InVivoModel (Xenograft)
  - caMOD::YeastModel
- Application specific classes under
  - caMOD::Comments & Logging
  - caMOD::Admin
  - caMOD::Preferences
- EVSTree utilizes caBIO 3.1 API





Vocabulary Usage

- NCI Thesaurus for murine organ and diagnosis terms
- NCI Thesaurus for human anatomical terms (release 2.1)
- NCI Thesaurus for staining methods (release 2.1)
- NCI Thesaurus for mouse and rat strains (release 2.1)
  - EVSTree – Reusable component for rendering EVS concepts



Common Data Elements