



- EVS Overview - Biomedical Terminology and Ontology Resources



Frank Hartel, Ph.D.

**Director, Enterprise Vocabulary Services
NCI Center for Bioinformatics**

EVS Goal – Integration by Meaning

- ◆ Clinical, translational, and basic research have overlapping but specialized needs
 - ◆ Inconsistent conceptual frameworks
 - ◆ Terminology and taxonomic conventions
 - May conflict
 - Evolve at different rates
- ◆ Knowledge model or terminology?
 - Tagging data: store/transfer/archive for future analysis
 - ◆ Reasoning: inference about data in a limited sense



Enterprise Vocabulary Services

- ◆ **Services and resources that address NCI's needs for controlled vocabulary** <http://ncicb.nci.nih.gov/core/EVS>
- ◆ **An NCI collaboration**
 - **NCI Office of Communications**
 - ◆ Cancer Information Products and Systems
 - ◆ PDQ and Cancer.gov
 - **NCI Center for Bioinformatics**
 - ◆ caCORE (built on EVS terminology)
 - ◆ Community portals



EVS - concluded

◆ Vocabulary Products

- NCI Thesaurus – an ontology-like terminology
- NCI Metathesaurus – maps vocabularies
- External vocabularies maintained and served

◆ Current Collaborations

- Federal Collaboration
- MGED Ontology
- Human Anatomy
- Cancer Classification
- MMHCC
- HL7, CHI
- caBIG Vocabulary/
Data Elements

NCI Thesaurus

- ◆ Reference Terminology for NCI
- ◆ Public domain, open content license
- ◆ Broad coverage of cancer domain
 - Neoplastic disease
 - Findings and Abnormalities
 - Anatomy
 - Agents, drugs, chemicals
 - Oncogenes, gene products, biological processes
 - Cancer models - murine
 - Research techniques, management

NCI Thesaurus - concluded

- ◆ Description-logic based (AL-)
- ◆ 34,000+ “Concepts” hierarchically organized
- ◆ 20 hierarchies, 19 “Kinds”
- ◆ “Roles” establish semantic relationships between Concepts
- ◆ “Properties” state facts about Concept
 - Links to LocusLink, OMIM, Swis, and so on
- ◆ Concept history

Structure of History Tables

TDE

Column Name	Description
History_ID	Record Number
Concept_Code	Concept Code
Concept_Name	Preferred Name of Concept
Action	Edit Action
Reference_Code	Referenced Concept Code
Edit_Date	Timestamp
Edit_Name	Name of edited NCI Thesaurus™ schema
Host	IP address of editor's workstation
Published	Publication state of history entry

DTS

Column Name	Description
History_ID	Record Number
Concept_Code	Concept Code
Action	Edit Action
Baseline_Date	Date of NCI Thesaurus™ Baseline
Reference_Code	Referenced Concept Code

NCI DTS Browser - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://ncievs-test2.nci.nih.gov:8080/NCIBrowser/Connect.do>

Links [Customize Links](#) [Free Hotmail](#) [Windows Media](#) [Windows](#) [NCICB-CBR Indexer](#) [Free AOL & Unlimited Internet](#)

Google Search Web Search Site PageRank Options

NATIONAL CANCER INSTITUTE

Go! **Root Concepts**

Go! **Quick Search**

Go! **Advanced Search**

Go! **History**

 Max Results:

Home Results About
 Help Customize Log out

Signaling Pathway Gene
 AKAP12 Gene
 AKAP13 Gene
 AMSH Gene
 APC Gene
 APCL Gene
 ARHGEF1 Gene
 ARHGEF5 Gene
 ARHH Gene
 ARRB1 Gene
 AXIN1 Gene
 AXIN2 Gene
 Adaptor Signaling Protein Gene
 BAG1 Gene
 BAG4 Gene
 BAG5 Gene
 BAI1 Gene
 BRD2 Gene
 BRDG1 Gene
 CAV1 Gene
 CAV2 Gene
 CBL Gene
 CCM1 Gene
 CNK1 Gene
 DAXX Gene
 DOK1 Gene
 EPS15 Gene
 EPS8 Gene
 G Protein Gene
 ARHA Gene
 ARHB Gene
 ARHC Gene
 ARHI Gene
 MAS1 Gene
 RRAD Gene
 RALA Gene
 RALB Gene
 RAS Family Gene
 HRAS Gene
 KRAS2 Gene

Concept Details

KRAS2 Gene [Generate URI](#)

Identifiers:

name	KRAS2 Gene
code	C25785

Roles:

Gene_Found_In_Organism	<input type="checkbox"/> <input checked="" type="checkbox"/> Human
Gene_In_Chromosomal_Location	<input type="checkbox"/> <input checked="" type="checkbox"/> 12p12.1
Gene_Plays_Role_in_Process	<input type="checkbox"/> <input checked="" type="checkbox"/> Signal Transduction

Properties:

Gene_Encodes_Product	p21 K-Ras Protein
Locus_ID	3845
OMIM_Number	190070
Preferred_Name	KRAS2 Gene
Semantic_Type	Gene or Genome
DEFINITION	NCI The KRAS2 gene encodes p21 K-RAS protein, two KRAS isoforms of which are produced by alternative splicing. RAS proteins are involved in transmembrane signal transduction as monomeric GTP-binding proteins, with intrinsic GTPase activity, that alternate between an inactive GDP-bound form and an active GTP-bound form. RAS proteins are activated by a guanine nucleotide-exchange factor and inactivated by a GTPase-activating protein. Mitogen-stimulated RAS enhances MYC accumulation by stabilizing MYC protein. This stabilization is dependent on the RAS/RAF/MAPK pathway, which appears to inhibit the proteasome-dependent degradation of MYC. Mutations of specific amino acids activate c-RAS to transform cells and are implicated in a variety of human tumors. KRAS is involved in malignancy much more often than is HRAS. (from OMIM, SWISS-PROT and NCI)
FULL_SYN	C-K-RAS SY NCI
FULL_SYN	K-RAS2A SY NCI
FULL_SYN	K-RAS2B SY NCI

NCI DTS Browser - Microsoft Internet Explorer

Address: http://ncievs-test2.nci.nih.gov:8080/NCIBrowser/Connect.do

Links: Customize Links, Free Hotmail, Windows Media, Windows, NCI-CBR Indexer, Free AOL & Unlimited Internet

Google Search Web Search Site PageRank Options

NATIONAL CANCER INSTITUTE

Go! Root Concepts

Go! Quick Search: c-Kit Gene

Go! Advanced Search: role, Anatomic_Structure_Has_Location

Go! History: Max Results 10

Home Help, Results Customize, About Log out

Signaling Pathway Gene

- . AKAP12 Gene
- . AKAP13 Gene
- . AMSH Gene
- . APC Gene
- . APCL Gene
- . ARHGEF1 Gene
- . ARHGEF5 Gene
- . ARHH Gene
- . ARRB1 Gene
- . AXIN1 Gene
- . AXIN2 Gene
- Adaptor Signaling Protein Gene
 - . BAG1 Gene
 - . BAG4 Gene
 - . BAG5 Gene
 - . BAI1 Gene
 - . BRD2 Gene
 - . BRDG1 Gene
 - . CAV1 Gene
 - . CAV2 Gene
 - . CBL Gene
 - . CCM1 Gene
 - . CNK1 Gene
 - . DAXX Gene
 - . DOK1 Gene
 - . EPS15 Gene
 - . EPS8 Gene
 - G Protein Gene
 - . ARHA Gene
 - . ARHB Gene
 - . ARHC Gene
 - . ARHI Gene
 - . MAS1 Gene
 - . RRAD Gene
 - . RALA Gene
 - . RALB Gene
 - RAS Family Gene
 - . HRAS Gene
 - . KRAS2 Gene

Concept Details

C-KIT Gene [Generate URI](#)

Identifiers:

name	C-KIT Gene
code	C18609

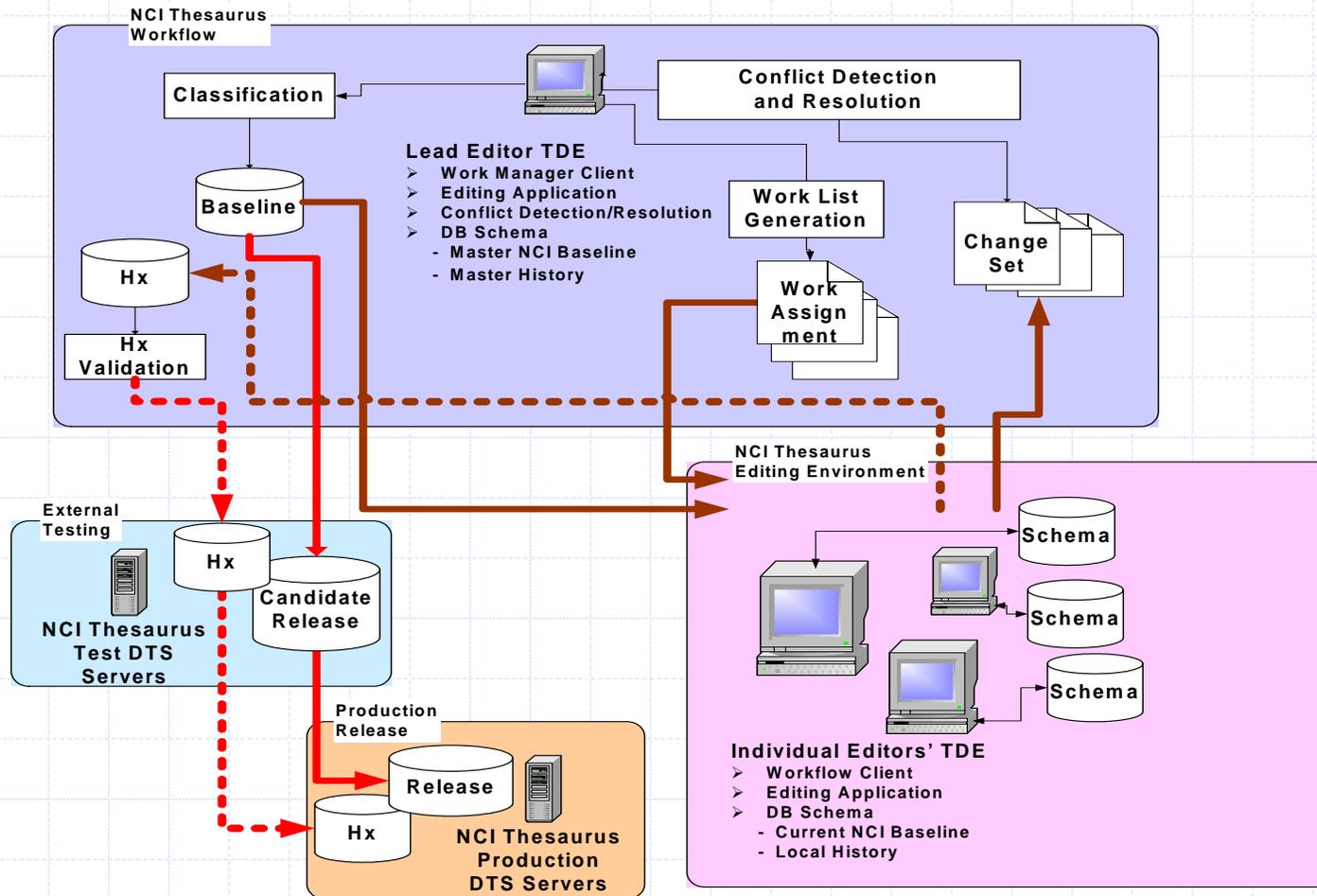
Roles:

Gene_Associated_With_Disease		Gastrointestinal Stromal Tumor
Gene_Found_In_Organism		Human
Gene_In_Chromosomal_Location		4q12
Gene_Plays_Role_in_Process		Gametogenesis
Gene_Plays_Role_in_Process		Hematopoiesis
Gene_Plays_Role_in_Process		Melanogenesis
Gene_Plays_Role_in_Process		Signal Transduction
Gene_Plays_Role_in_Process		Tyrosine Phosphorylation

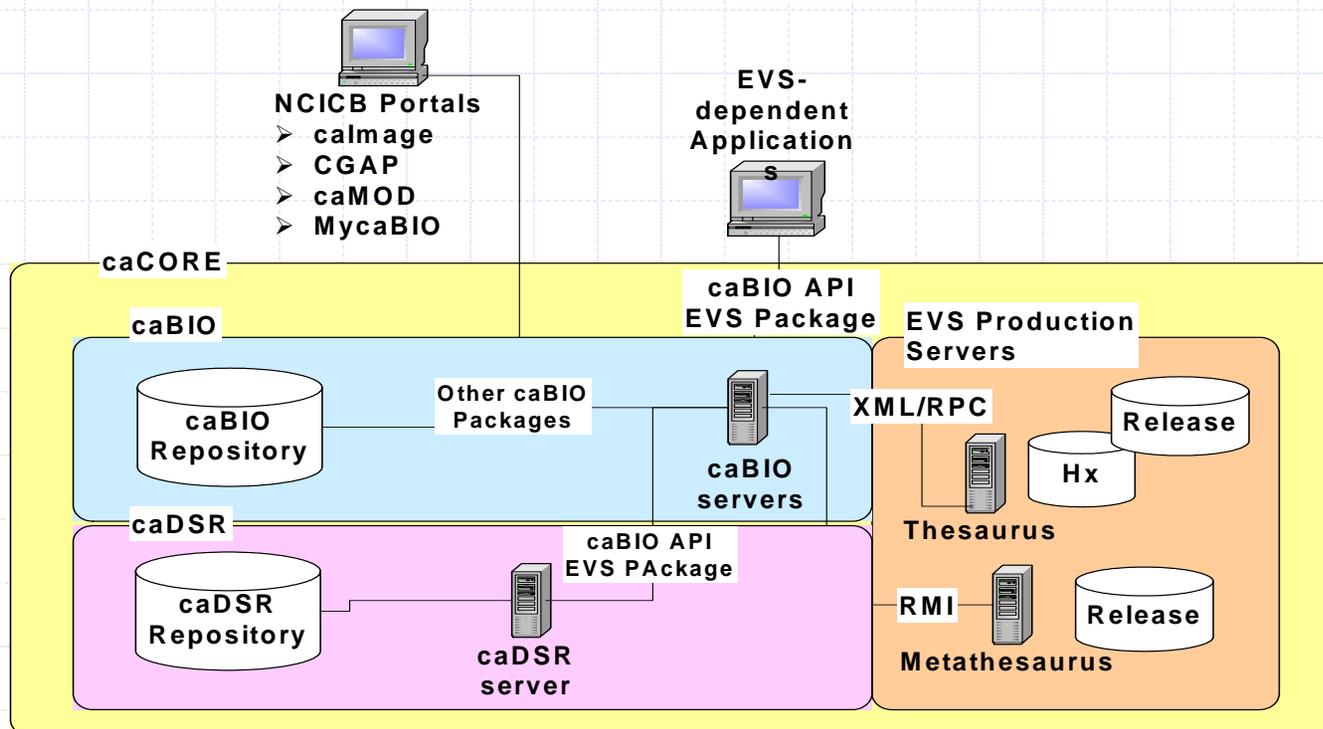
Properties:

GenBank_Accession_Number	X06182
Gene_Encodes_Product	Stem Cell Factor Receptor Tyrosine Protein Kinase
Homologous_Gene	CSF1R/PDGFR Family
Locus_ID	3815
OMIM_Number	164920
Preferred_Name	C-KIT Gene
Semantic_Type	Gene or Genome
Swiss_Prot	P10721

NCI Thesaurus Production Environment

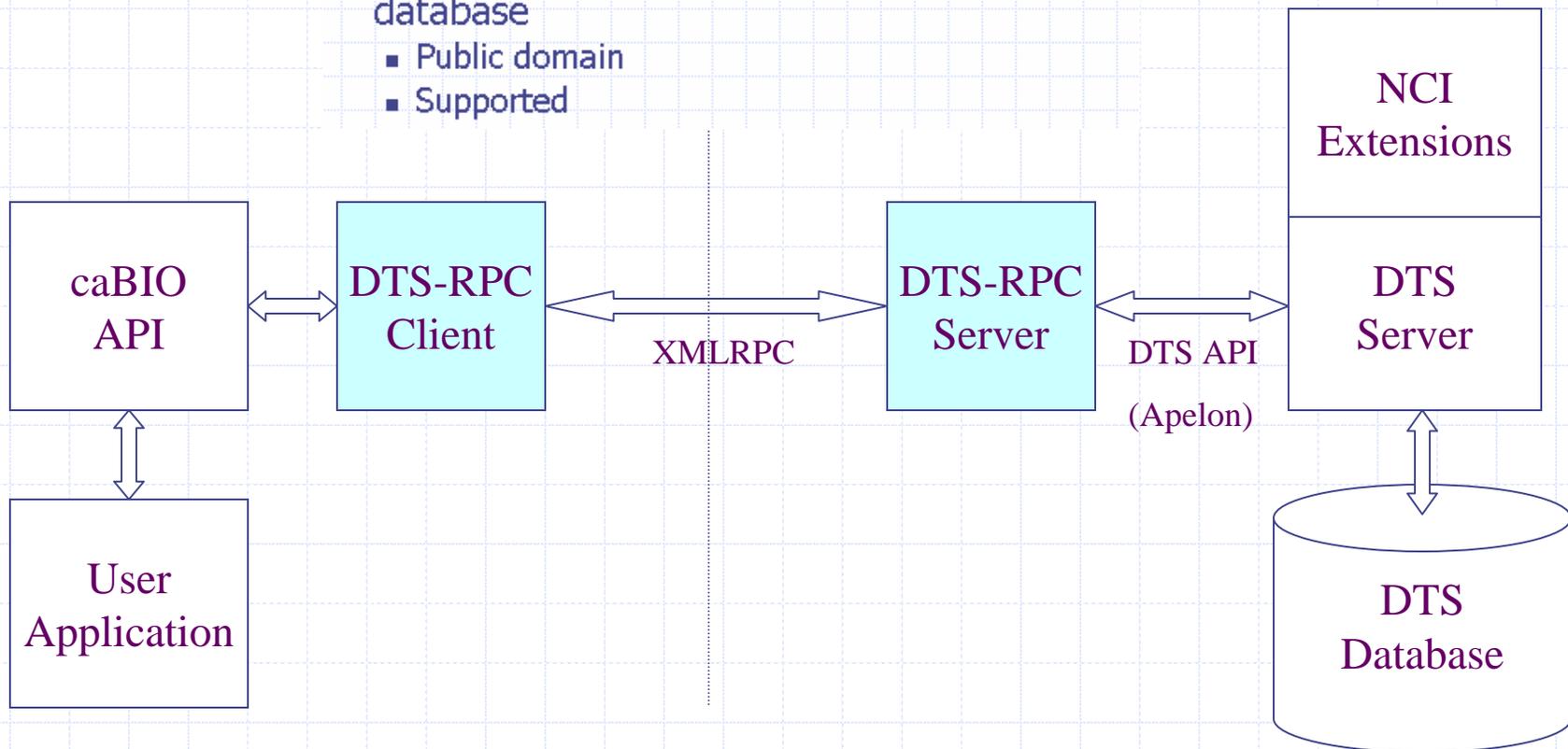


NCICB builds on EVS via caCORE Infrastructure



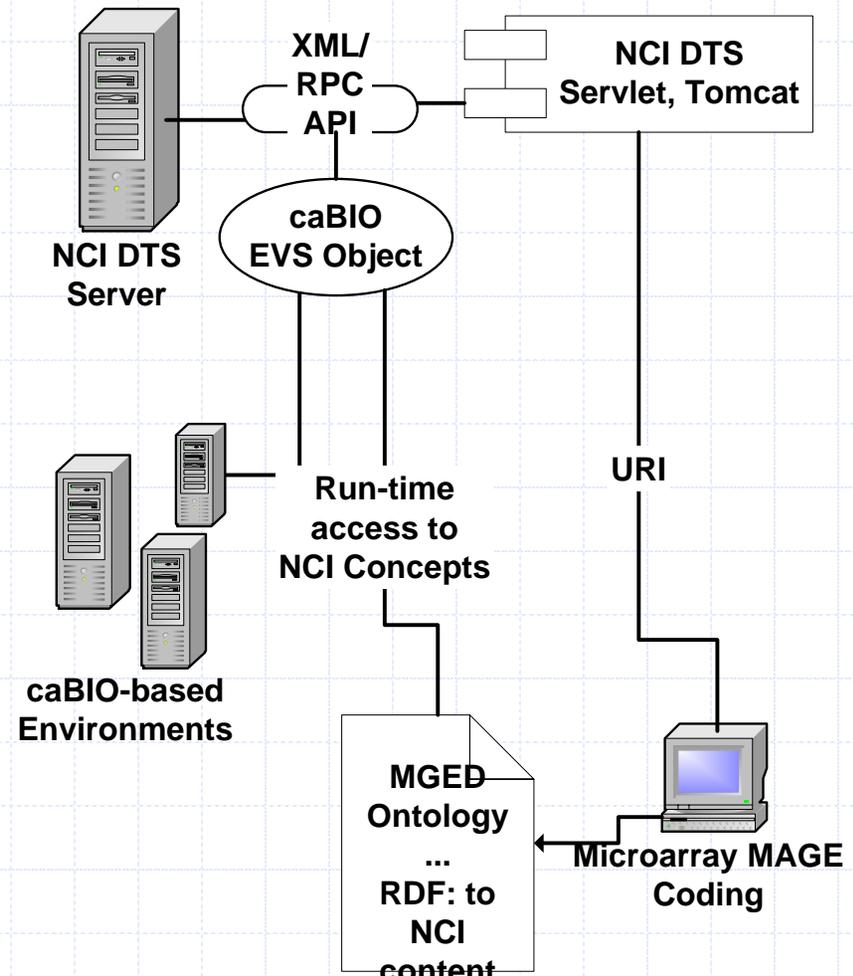
Server API extensions (DTSRPC)

- ◆ Provide capability for the users to retrieve vocabulary and edit history data from DTS database
 - Public domain
 - Supported

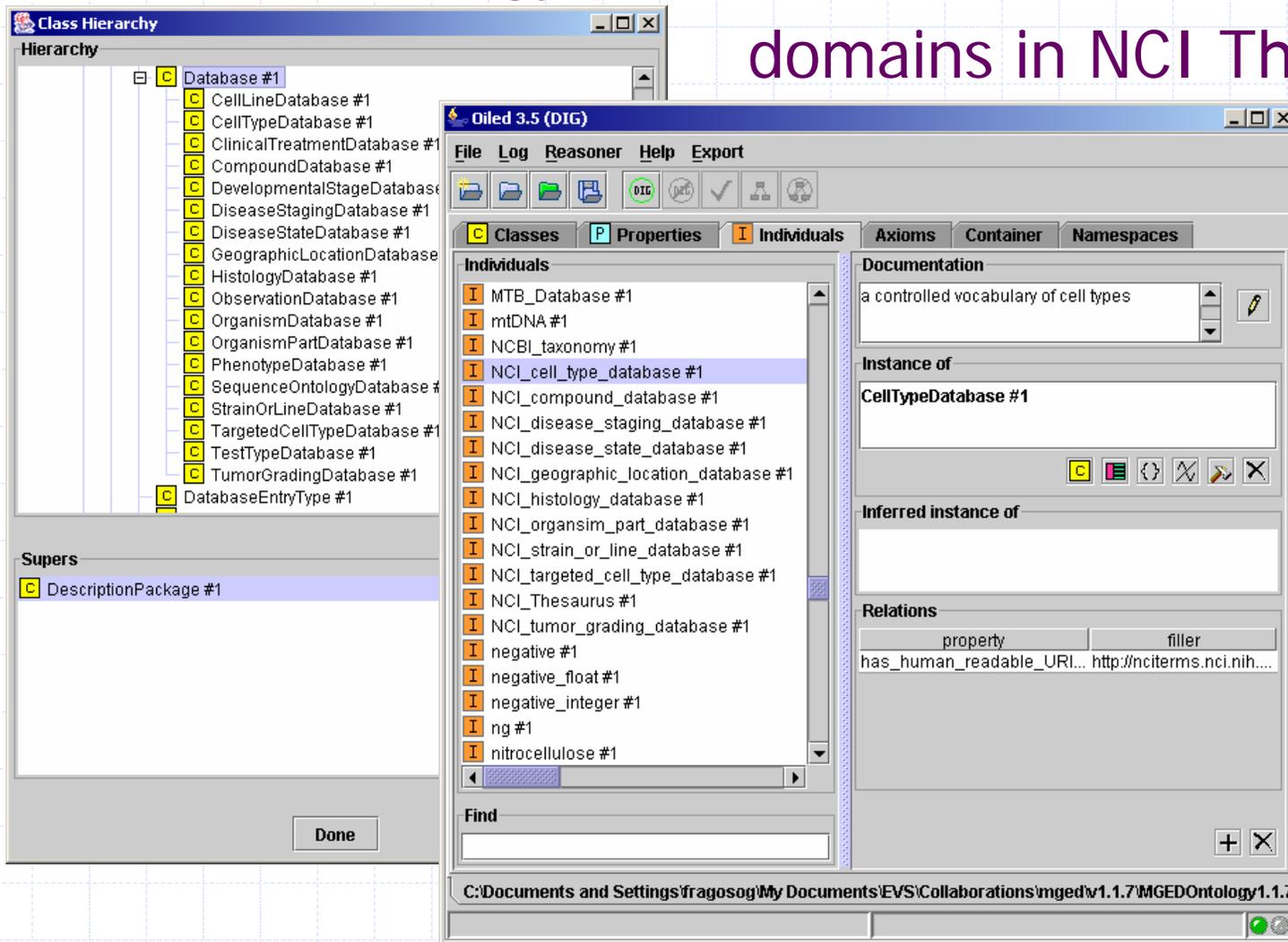


NCI Thesaurus access

- ◆ **MGED Ontology uses DAML+OIL**
 - **daml+oil allows inclusion of external ontology content via RDF**
 - **NCI DTS server, servlet enhanced with URI**
 - **Enables MGED Ontology to specify NCI content via reference**



MGED Ontology points to various domains in NCI Thesaurus



Class Hierarchy

- Database #1
 - CellLineDatabase #1
 - CellTypeDatabase #1
 - ClinicalTreatmentDatabase #1
 - CompoundDatabase #1
 - DevelopmentalStageDatabase #1
 - DiseaseStagingDatabase #1
 - DiseaseStateDatabase #1
 - GeographicLocationDatabase #1
 - HistologyDatabase #1
 - ObservationDatabase #1
 - OrganismDatabase #1
 - OrganismPartDatabase #1
 - PhenotypeDatabase #1
 - SequenceOntologyDatabase #1
 - StrainOrLineDatabase #1
 - TargetedCellTypeDatabase #1
 - TestTypeDatabase #1
 - TumorGradingDatabase #1
- DatabaseEntryType #1

Supers

- DescriptionPackage #1

Oiled 3.5 (DIG)

File Log Reasoner Help Export

Classes Properties Individuals Axioms Container Namespaces

Individuals

- MTB_Database #1
- mtDNA #1
- NCBI_taxonomy #1
- NCI_cell_type_database #1
- NCI_compound_database #1
- NCI_disease_staging_database #1
- NCI_disease_state_database #1
- NCI_geographic_location_database #1
- NCI_histology_database #1
- NCI_organsim_part_database #1
- NCI_strain_or_line_database #1
- NCI_targeted_cell_type_database #1
- NCI_Thesaurus #1
- NCI_tumor_grading_database #1
- negative #1
- negative_float #1
- negative_integer #1
- ng #1
- nitrocellulose #1

Documentation

a controlled vocabulary of cell types

Instance of

CellTypeDatabase #1

Inferred instance of

Relations

property	filler
has_human_readable_URI...	http://nciterns.nci.nih...

Find

C:\Documents and Settings\fragosog\My Documents\EVS\Collaborations\mged\v1.1.7\MGEDOntology1.1.7

Cell Types
 Drugs / Chemicals
 Neoplastic Diseases
 Staging
 Tumor Grading
 Histology
 Organism Parts
 Strains, Lines
 Geographic Location

◆ Biomaterials

- Cell Type
- Organism_part

Encoding NCI content in MAGE-ML

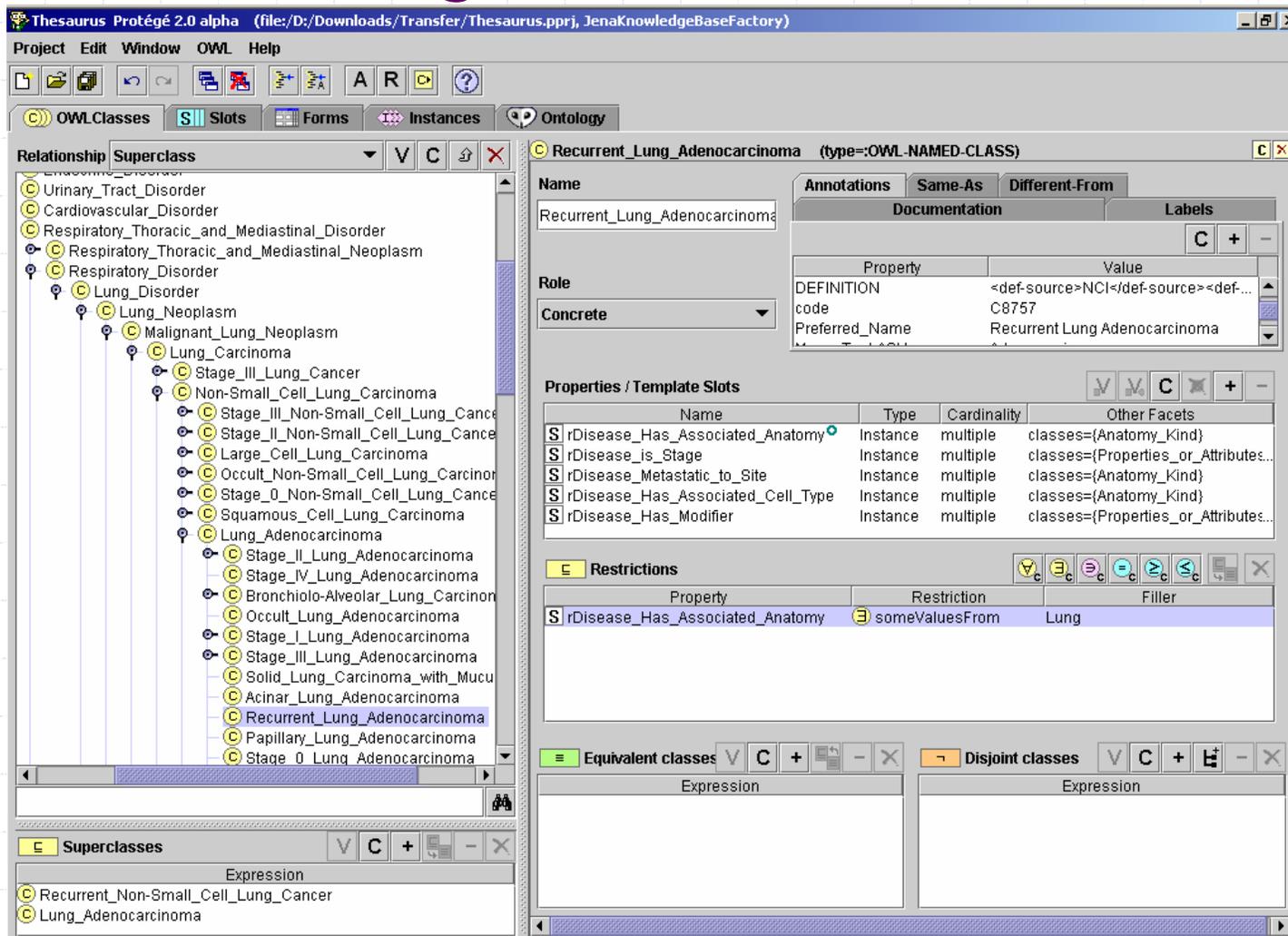
```
...  
<!-- The cell type of AD145: Epithelial Cell.
```

```
    The term was obtained from the NCI Thesaurus, has "Somatic Cell" as parent  
    concept. For purpose of example coding only, prefer using MO:CellType.
```

```
-->  
  <OntologyEntry category="Somatic Cell" value="Epithelial Cell">  
    <OntologyReference_assn>  
      <DatabaseEntry accession="C12578"  
URI="http://nciterns.nci.nih.gov/NCIBrowser/ConceptReport.jsp?dictionary=  
NCI+Thesaurus&#38;code=C12578">  
        <Database_assnref>  
          <Database_ref identifier="DB:nci_thesaurus" />  
        </Database_assnref>  
      </DatabaseEntry>  
    </OntologyReference_assn>  
  </OntologyEntry>
```

```
...
```

Future: Protégé/OWL ?



The screenshot shows the Protégé 2.0 alpha interface. The left pane displays a hierarchical tree of classes under 'Relationship Superclass'. The right pane shows the details for the selected class, 'Recurrent_Lung_Adenocarcinoma'.

Class Details: Recurrent_Lung_Adenocarcinoma (type=OWL-NAMED-CLASS)

Property	Value
DEFINITION	<def-source>NCI</def-source><def... code C8757
Preferred_Name	Recurrent Lung Adenocarcinoma

Name	Type	Cardinality	Other Facets
S rDisease_Has_Associated_Anatomy	Instance	multiple	classes={Anatomy_Kind}
S rDisease_Is_Stage	Instance	multiple	classes={Properties_or_Attributes...}
S rDisease_Metastatic_to_Site	Instance	multiple	classes={Anatomy_Kind}
S rDisease_Has_Associated_Cell_Type	Instance	multiple	classes={Anatomy_Kind}
S rDisease_Has_Modifier	Instance	multiple	classes={Properties_or_Attributes...}

Property	Restriction	Filler
S rDisease_Has_Associated_Anatomy	someValuesFrom	Lung

At the bottom, there are sections for 'Equivalent classes' and 'Disjoint classes', both currently empty.

NCI Metathesaurus

- ◆ **UMLS Metathesaurus extended with cancer-oriented vocabularies**
 - **780,000+ concepts, 1,700,000 terms and phrases**
 - **Mappings among over 50 vocabularies**
 - **Rich synonymy: Over 40,000 terms for “cancer” mapped to 7,000 concepts**
- ◆ **Used as online dictionary, thesaurus, for mapping and document indexing**
- ◆ **Accessible via caBIO**

(q22;q12)(PML/RARa)

Search

The basic search is enabled. Click on the "Advanced" tab to customize your search criteria. You can mouse-over each advanced search item for help in utilizing the advanced features.

[Concept](#) | [Definitions](#) | [Synonyms](#) | [Sources](#) | [Broader Concepts](#) | [Narrower Concepts](#) | [Related Concepts](#) | [Medications](#) | [Procedures](#) | [Laboratory](#) | [Diagnosis](#) | [Open NCI Hierarchy](#) | [View Hierarchy Location](#)

CL106989: Acute Promyelocytic Leukemia with t(15;17)(q22;q12)(PML/RARa)

Neoplastic Process

[Acute Promyelocytic Leukemia with t\(15;17\)\(q22;q12\)\(PML/RARa\) Definitions](#)

Source	Definition
ICDO3	includes all variants
NCI-GLOSS	A type of acute myeloid leukemia, a quickly progressing disease in which too many immature blood-forming cells are found in the blood and bone marrow.
NCI	An acute myeloid leukemia (AML) in which abnormal promyelocytes predominate. It is characterized by the t(15;17)(q22;q12) translocation. There are two variants: the typical and micro granular (hypo granular) variant. This AML is particularly sensitive to treatment with all trans-retinoic acid and has a favorable prognosis. (WHO, 2001) -- 2003
NCI	Maps_To_LASH: leukemia

[Acute Promyelocytic Leukemia with t\(15;17\)\(q22;q12\)\(PML/RARa\) Synonyms](#)

- Acute Promyelocytic Leukemia with t(15;17)(q22;q12)(PML/RARa)
- Acute Promyelocytic Leukemia with t(15;17)(q22;q12)
- Acute promyelocytic leukemia, t(15;17)(q22;q11-12)
- Acute myeloid leukemia, t(15;17)(q22;q11-12)

[Acute Promyelocytic Leukemia with t\(15;17\)\(q22;q12\)\(PML/RARa\) Sources](#)

NCI ICDO3

View neighborhood in

Table View Of [Sources](#)

[Concepts More General Than Acute Promyelocytic Leukemia with t\(15;17\)\(q22;q12\)\(PML/RARa\)](#)

- About
- Browse
- Copyright
- New Term
- Sources
- User's Guide

Center
for
Bioinformatics

FIRSTGOV
Year First Click to the U.S. Government

- Metabolic Disorder
- Neoplasm
 - Neoplasm by Morphology
 - Neoplasm by Site
 - Neoplasm by Special Category
 - Benign Neoplasm
 - Childhood Neoplasm
 - Common Neoplasm
 - Astrocytic Tumor
 - Common Carcinoma
 - Common Connective and Soft Tissue Neoplasms
 - Common Germ Cell Neoplasm
 - Common Hematopoietic Neoplasm
 - Acute Myeloid Leukemia
 - Acute Eosinophilic Leukemia
 - Acute Leukemia of Ambiguous Lineage
 - Acute Myeloid Leukemia in Remission
 - Acute Myeloid Leukemia not Otherwise Specified
 - Acute Myeloid Leukemia with Multilineage Potential
 - Acute Myeloid Leukemia with Recurrent Genetic Abnormalities
 - Acute Myeloid Leukemia with 11q23 Abnormality
 - Acute Myeloid Leukemia with Abnormalities of Chromosomes 5, 7, and 17
 - Acute Myeloid Leukemia with t(8;21)(q22;q22)
 - Acute Promyelocytic Leukemia
 - Acute Promyelocytic Leukemia
 - Adult Acute Promyelocytic Leukemia

Publication Cycle

◆ NCI Thesaurus

- Monthly
- History applies to published concepts
- Formats
 - ◆ Ontylog XML
 - ◆ OWL
 - ◆ Flat file

◆ NCI Metathesaurus

- Minor releases monthly
- Major releases twice a year
- Format
 - ◆ MR+

EVS Team

EVS

NCI OC – oncology, pathology, pharmacy

Margaret Haber

Larry Wright

NCI CB – biology, operations

Sherri Coronado

Gilberto Fragoso

Frank Hartel

Apelon, Inc.

Aspen, Inc.

Jim Oberthaler Consulting

Northrup Grumman, Inc.

Kevric Corporation



NCI DTS Browser - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites History

Address <http://nciterns.nci.nih.gov/NCIBrowser/Connect.do> Go



Home Results About
Help Customize Log out

Go! Root Concepts

Go! Quick Search

Go! Advanced Search

Go! History

Max Results

- [-] Pulmonary Sulcus Neoplasm
- [-] Malignant Respiratory Tract Neoplasm
- [-] Malignant Lung Neoplasm
 - [+] Lung Blastoma
 - [-] Lung Carcinoma
 - . Asbestos-Related Lung Carcinoma
 - [+] Combined Lung Carcinoma
 - . Lung Adenoid Cystic Carcinoma
 - . Lung Hilum Carcinoma
 - . Metastatic Carcinoma to the Lung
 - [-] Non-Small Cell Lung Carcinoma
 - [+] Adenosquamous Lung Carcinoma
 - [+] Large Cell Lung Carcinoma
 - [-] Lung Adenocarcinoma
 - . Acinar Lung Adenocarcinoma
 - [+] Bronchiolo-Alveolar Lung Carcinoma
 - . Bronchogenic Lung Adenocarcinoma
 - . Occult Lung Adenocarcinoma
 - . Papillary Lung Adenocarcinoma
 - . **Recurrent Lung Adenocarcinoma**
 - . Solid Lung Carcinoma with Mucus Formation
 - . Stage 0 Lung Adenocarcinoma
 - [+] Stage I Lung Adenocarcinoma
 - [+] Stage II Lung Adenocarcinoma
 - [+] Stage III Lung Adenocarcinoma
 - . Stage IV Lung Adenocarcinoma

Concept Details

Recurrent Lung Adenocarcinoma

[Generate URI](#)

Identifiers:

name	Recurrent Lung Adenocarcinoma
code	C8757

Roles:

Disease_Has_Associated_Anatomy		Lung
Disease_Has_Associated_Anatomy		Respiratory System

Properties:

Maps_To_LASH	Adenocarcinoma
Preferred_Name	Recurrent Lung Adenocarcinoma
Semantic_Type	Neoplastic Process
TEMP_CUI	CL032662
DEFINITION	NCI The reemergence of adenocarcinoma cell lung cancer after a

Done Internet