caARRAY Interoperability Review Presentation to V/CDE Workspace

Presenter: Juergen Lorenz (SAIC)

Date: 12/01/2005

NATIONAL INSTITUTES OF HEALTH
DEPARTMENT OF HEALTH & HUMAN SERVICES • USA
SAIC From Science to Solutions™
NATIONAL CANCER INSTITUTE®
Center for Bioinformatics

caARRAY

- Mervi Heiskanen
- Anand Basu
- Xiaopeng Bian
- Jerry Eads
- Jim Zhou
- Hangjiong Chen
- Sumeet Muju
- Phu Tran
- Juergen Lorenz
- Scott Gustafson
- Himanso Sahni
- Durga Addepalli
- Ye Wu
- Don Swan
- Kevin Fitzpatrick
- John Moy
- Sharon Settnek
- Tara Akhavan

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caARRAY

- microarray database system which has been designed, architected and constructed to be compliant with international standardization efforts in the field of microarray and gene expression-related information persistence and processing

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caARRAY

- MAGE
  - MicroArray and GeneExpression Object Model and Markup Language
  - 1.1 (October 2003)
  - http://www.omg.org/docs/formal/03-10-01.pdf
- MIAME
  - Minimum Information About a Microarray Experiment
  - 1.1 Draft 6 (April 1, 2002)
  - http://www.mged.org/Workgroups/MIAME/miame_1.1.html

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caARRAY

- strives for Interoperability (syntactically and semantically) in compliance with:
  - MAGE
  - MIAME
  - caBIG

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caARRAY

- microarray database system which has been designed, architected and constructed to facilitate interoperability with
  - MAGE information providers (e.g. other microarray database systems),
  - MAGE information consumers (e.g microarray analysis software)

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caARRAY

- Interoperability with other MAGE information providers and consumers is achieved thru well-documented interfaces:
  - MAGE-OM API (programmatically)
    - Provides fine grain search and retrieval of MAGE information from caARRAY via a caBIO-like API
  - MAGE-ML (document-based)
    - MAGE-ML import/export

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caARRAY

- strives to be interoperable[1] with other MAGE information providers[2]:
  - MAGE Databases/Portals
    - ArrayExpress (EBI)[1]
    - others
  - Microarray imaging software[3]
    - GenePix (Molecular Devices)[3]
    - GCOS (Affymetrix)[1, 3]
    - Imagene (BioDiscovery)[3]
    - others

[1] document based via MAGE-ML [2] outside caBIG [3] support for native file formats

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caARRAY

- strives to be interoperable[1] with other MAGE information consumers[2]:
  - Microarray analysis software
    - caWorkBench (Columbia University)[2]
    - webGenome (RTI)[2]
    - others

[1] programmatically via MAGE-OM API [2] in caBIG

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caWorkBench

caWorkBench2.0 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://amdec-bioinfo.cu-genome.org/html/caWorkBench.htm Go

MDeC Bioinformatics Core Facility
at the Columbia Genome Center

HOME | AMDeC | Columbia Genome Center | Contact

About The Center
- Introduction
- What you can do
- Using the Facility
- Hardware
- Software
- Databases
- Staff

Services for Users
- Access
- Manuals
- Support
- Registration

Resources
- caWorkBench3.0
- Algorithm Reference
- Tutorials & Examples
- Links
- Maps & Directions
- Contact Us

caWorkBench2.0 - A Platform for Integrated Genomics

Introduction	Features	Installation Requirements	Application Download	Sample Data	Documentation and Support	Video Tutorials

Introduction

caWorkBench2.0, an open source bioinformatics platform written in Java, makes sophisticated tools for data management, analysis and visualization available to the community in a convenient fashion. It evolved from a project which was originally sponsored by the National Cancer Institute Center for Bioinformatics (NCICB). Some of the most fully developed capabilities of the platform include microarray data analysis, pathway analysis and reverse engineering, sequence analysis, transcription factor binding site analysis, and pattern discovery.

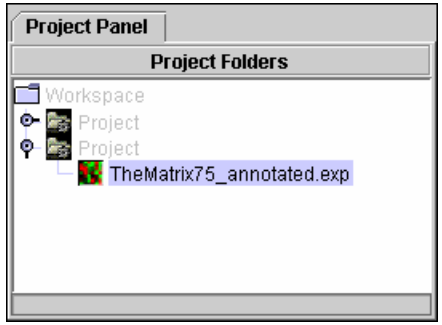
The focus in caWorkBench2.0 is on data and algorithm integration. As a result, caWorkBench2.0 users will be naturally able to combine the analysis in over 50 available caWorkBench2.0 components using a

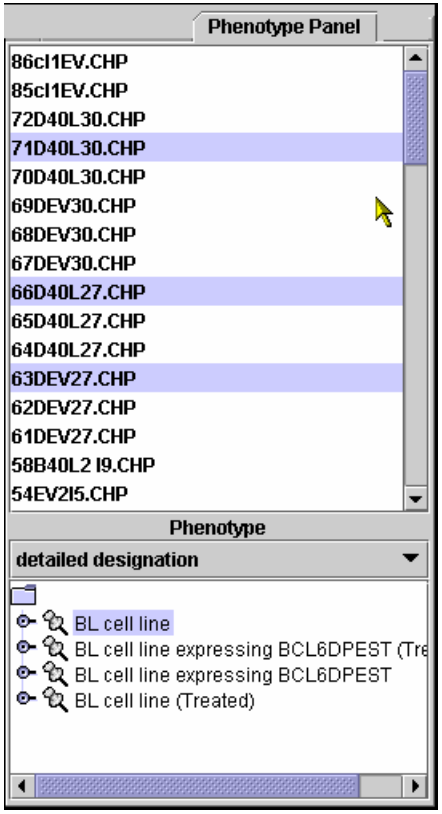
Internet

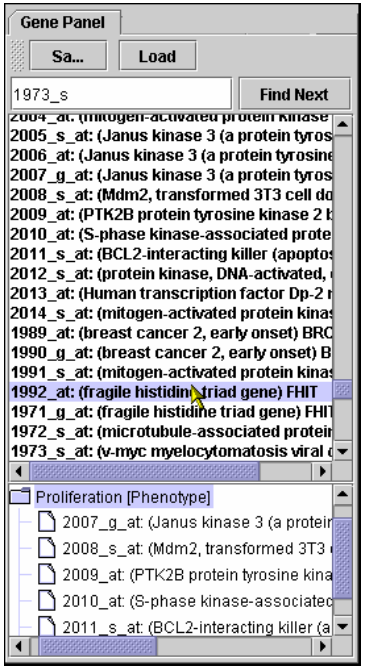
caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caWorkBench

Data Management







caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caWorkBench

Color Mosaic

Pat Abs All Ratio Acce... Label ☑ All arrays ☑ All Markers

Go Term Panel

Process Sa... Load

behavior(15)
- adult behavior(0)
- behavioral fear response(0)
- behavioral response to water deprivation(0)
- chemosensory behavior(2)
- feeding behavior(17)
- grooming behavior(0)
- larval behavior(0)
- learning and/or memory(11)
- locomotory behavior(5)
- mechanosensory behavior(2)
- reproductive behavior(0)
- rhythmic behavior(3)

feeding behavior [GO Term]
- 2039_s_at: (FYN oncogene related to SRC, FGR, YES) FYN
- 31417_at: (hypocretin (orexin) neuropeptide precursor) HCR
- 31905_at: (galanin receptor 2) GALR2
- 32998_at: (cholecystokinin A receptor) CCKAR
- 33070_at: (hypocretin (orexin) receptor 1) HCRTR1
- 33072_at: (hypocretin (orexin) receptor 2) HCRTR2
- 33525_at: (neuropeptide Y receptor Y2) NPY2R
- 33550_at: (5-hydroxytryptamine (serotonin) receptor 2C) HT
- 33551_s_at: (5-hydroxytryptamine (serotonin) receptor 2C)
- 34597_at: (pancreatic polypeptide receptor 1) PPYR1
- 35068_at: (agouti related protein homolog (mouse)) AGRP
- 35437_at: (neuropeptide Y receptor Y1) NPY1R

Dendrogram

☑ Enable Zoom 8 6

Reverse Engineering | Analysis

Profiler | Visualization | Conditional Analysis | Genetic Network | jPanel5

Layout GEM Clear Set Parameters Print Load Draw SF Histo Gene: 3_s_at Depth: 2

Name: v-myc myelocytomatosis viral oncogene homolog (avian) Affymetrix Id: 1827_s_at

Function	Process
DNA binding, -- transcription factor activity, -- nucleus, -- regulation of transcription, DNA-dependent, -- regulation of transcription from Pol II promoter, -- iron ion homeostasis, -- cell cycle arrest, -- cell growth and/or maintenance, -- cell proliferation,	Erk1/Erk2 MAPK Signaling Pathway; -- p38 MAPK Signaling Pathway; -- Overview of telomerase protein component gene hTert

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

webGenome

webGenome: Select Experiments - Microsoft Internet Explorer

File Edit View Favorites Tools Help

webGenome

dhall logged in Logout Help

Overview | Plot Data | Virtual Experiments | Analytic Pipelines

Data | Plot Parameters | Scatter Plot | Annotation Plot | Annotation Report | Ideogram Plot | Probe Plot

Select Experiments → Select Experiment Type → Select Probe Sets → Select Assembly → Mapping Completed

Select	Experiment Name	Database	Experiment Description	Contact
☐	experiment_1	Excel	Test experiment 1	dhall@rti.org
☐	experiment_2	Excel	Test experiment 2	dhall@rti.org
☐	expression_data	Excel	Test experiment 1	dhall@rti.org
☐	small_experiment	Excel	Test experiment	dhall@rti.org
☐	Submitted by agarwal	SKY/M-FISH&CGH	SKY/M-FISH&CGH experiments submitted by agarwal	agarwal
☐	Submitted by ajschaeffer	SKY/M-FISH&CGH	SKY/M-FISH&CGH experiments submitted by ajschaeffer	ajschaeffer
☐	Submitted by alfhm	SKY/M-FISH&CGH	SKY/M-FISH&CGH experiments submitted by alfhm	alfhm
☐	Submitted by difilip	SKY/M-FISH&CGH	SKY/M-FISH&CGH experiments submitted by difilip	difilip
☐	Submitted by ghadimi	SKY/M-	SKY/M-FISH&CGH experiments submitted by	ghadimi

Done | Local intranet

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

```
// Initialize secure session with MAGE-OM server
SecureSession session = SecureSessionFactory.defaultSecureSession
session.direct(sessionManagerUrl);
session.start(userName, password);

// Set up search criteria
ExperimentSearchCriteria searchCriteria =
   SearchCriteriaFactory.new_EXPERIMENT_EXPERIMENT_SC();
searchCriteria.setSessionId(session);

// Perform search
SearchResult result = searchCriteria.search();
Experiment[] experiments = (Experiment[])result.getResultSet();
```


caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

webGenome

webGenome: Scatter Plot - Microsoft Internet Explorer

File Edit View Favorites Tools Help

webGenome

dhall logged in Logout Help

Overview | Plot Data | Virtual Experiments | Analytic Pipelines

Data | Plot Parameters | Scatter Plot | Annotation Plot | Annotation Report | Ideogram Plot | Probe Plot

Show: Lines Data points

Fold change across chromosome 1

log2 Ratio

2, 1, 0, -1, -2, -3

bin_146

0, 50, 100, 150, 200

Chromosome Position (MB)

UOK143, UOK122, UOK121, UOK181, UOK115, UOK117, UOK154, UOK171, UOK111, UOK130

Start: mb End: mb Min ratio: Max ratio: Zoom

Local intranet

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caARRAY

- allows day to day management and analysis of microarray data, and facilitates data sharing between research centers
- designed to capture MIAME 1.1 level annotations
- supports programmatic query and retrieval of gene expression data via MAGE-OM API
- supports data exchange complying with international standardized format MAGE-ML
- data can be migrated to a central caARRAY database at the NCI or caARRAY can be deployed locally at other affiliated organizations
- supports submission and retrieval of Affymetrix, GenePix, and ImaGene native format data files

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caARRAY

Search Protocols - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://caarraydb.nci.nih.gov/caarray/searchProtocols.do?mode=setup Go

cancer.gov

caARRAY

NATIONAL CANCER INSTITUTE

PROTOCOLS

HOME | CONTACTS | PROTOCOLS | ARRAYS | ONTOLOGIES | BIOMATERIALS | EXPERIMENTS | LOGOUT

- SEARCH HARDWARE
- NEW HARDWARE
- SEARCH SOFTWARE
- NEW SOFTWARE
- SEARCH PROTOCOLS
- NEW PROTOCOL

QUICK LINKS

- NCI HOME
- NCICB HOME
- caARRAY WEBSITE
- NCICB SUPPORT
- HELP
- ABOUT

SEARCH PROTOCOLS

Protocol Type	All
Protocol Name	
Visibility	All

Search Reset Help

CONTACT US | PRIVACY NOTICE | DISCLAIMER | ACCESSIBILITY | APPLICATION SUPPORT

Internet

caBIG™ cancer Biomedical Informatics Grid™

NATIONAL INSTITUTES OF HEALTH

NATIONAL CANCER INSTITUTE

caARRAY

Hybridization File Upload Hybridization Protocol - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://caarraydb.nci.nih.gov/caarray/hybridizationHybProtocol.do

cancer.gov

caARRAY

NATIONAL CANCER INSTITUTE

EXPERIMENT | HOME | CONTACTS | PROTOCOLS | ARRAYS | ONTOLOGIES | BIOMATERIALS | EXPERIMENTS | LOGOUT

- IMPORT MAGE-ML
- SEARCH/MODIFY
- NEW EXPERIMENT

QUICK LINKS

- NCI HOME
- NCICB HOME
- caARRAY WEBSITE
- NCICB SUPPORT
- HELP
- ABOUT

COMMON HYBRIDIZATION INFORMATION	
Array Design:	hg_u95av2
Number of Channels:	1
Hybridization File Type:	Affymetrix
Array Group Template Name:	U95aV2-1_3

Hybridization Protocol: Affymatrix hybridization u95av2

If the protocol you would like to select is not listed, you can add this protocol to the caArray database (How?).

* indicates a required field

PROTOCOL	
Protocol ID:	gov.nih.nci.ncicb.caarray:Protocol:1015897549070016:1
* Activity Date:	
* Performer:	
Protocol Name:	Affymatrix hybridization u95av2
Protocol Description:	
URI:	
Protocol Parameters:	NAME / VALUE / TYPE

Next Cancel Help

CONTACT US | PRIVACY NOTICE | DISCLAIMER | ACCESSIBILITY | APPLICATION SUPPORT

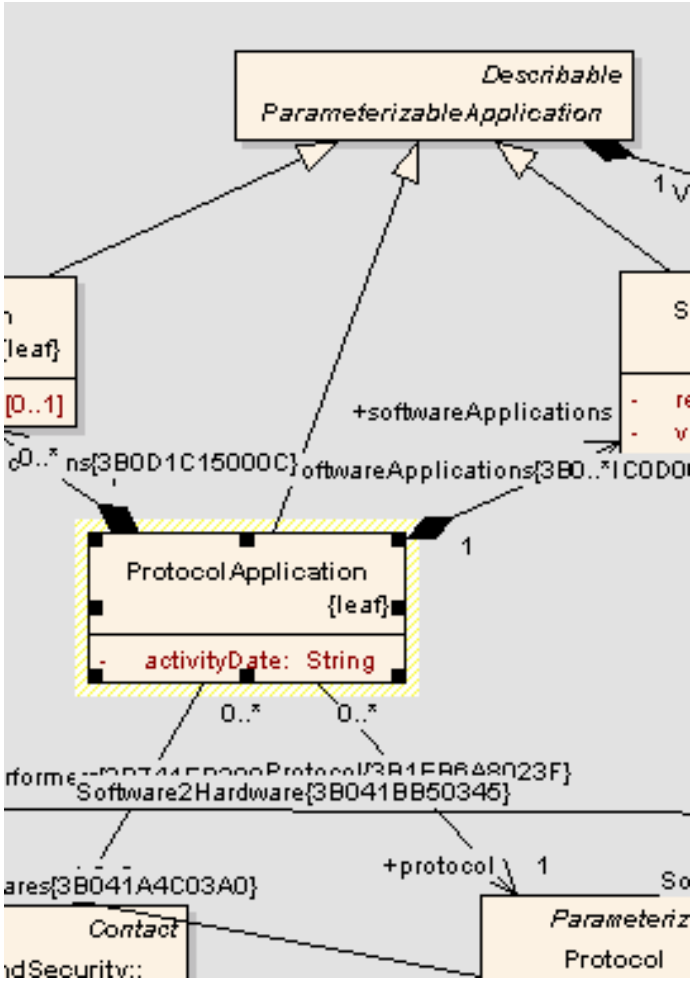
cancer.gov National Institutes of Health (NIH) HEALTH & HUMAN SERVICES FIRSTGOV Your First Click to the U.S. Government

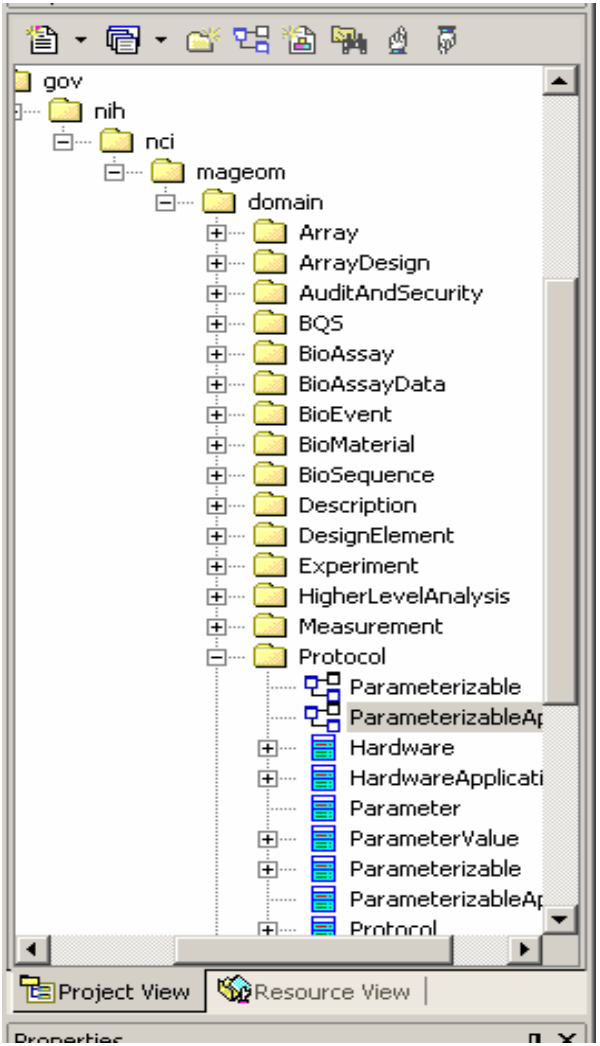
Internet

caBIG™ cancer Biomedical Informatics Grid™

NATIONAL INSTITUTES OF HEALTH NATIONAL CANCER INSTITUTE

caARRAY





caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caARRAY

Data Element Concept - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://cdebrowser.nci.nih.gov/CDEBrowser/search?dataElementConceptDetails=9&tabClicked=1&PageId=GetDetailsGroup Go Links

Public ID:	2236762
Version:	1.0
Long Name:	Protocol Application
Preferred Name:	C44184
Context:	caBIG
Qualifier:	

Object Class Concepts

Concept Name	Concept Code	Public ID	Definition Source	EVS Source	Primary
Protocol Application	C44184	2236570	NCICB	NCI_CONCEPT_CODE	Yes

Property

Public ID:	2236660
Version:	1.0
Long Name:	Activity Date
Preferred Name:	C43431:C25164
Context:	caBIG
Qualifier:	

Property Concepts

Concept Name	Concept Code	Public ID	Definition Source	EVS Source	Primary
Activity	C43431	2222740	NCI	NCI_CONCEPT_CODE	No
Date	C25164	2202880	NCI	NCI_CONCEPT_CODE	Yes

User: Public User Version 3.0.1.2 Build 20 Please send comments and suggestions to ncicb@pop.nci.nih.gov

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caARRAY

NCI Terminology Browser - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://nciterms.nci.nih.gov/NCIBrowser/Connect.do?bookmarktag=1&dictionary=NCI_Thesaurus&&code=C43431

National Cancer Institute — U.S. National Institutes of Health | www.cancer.gov

Quick Search

Max Results: 100

Advanced Search: Information about this concept — Accepted_Therapeutic_Use_For

Vocabulary: NCI_Thesaurus

Concepts visited (during this session): Activity

QUICK LINKS: NCI HOME, NCICB HOME, EVS HOME, SITE HOME, KNOWN ISSUES

HELP | RESULTS | CUSTOMIZE | ABOUT | BROWSE HIERARCHY | LOGOUT

Concept Details

Bookmark this page

Activity — Printable Page | History | Graph

Identifiers:

name	Activity
code	C43431

Information about this concept:

DEFINITION	NCI|The state of being active; any specific activity.
Synonym with source data	Activity|PT|NCI
Preferred_Name	Activity
Semantic_Type	Activity
Synonym	Activity

Superconcepts

Conceptual_Entities

Error on page. — Trusted sites

caBIG™ cancer Biomedical Informatics Grid™ NATIONAL INSTITUTES OF HEALTH NATIONAL CANCER INSTITUTE

caARRAY

NCI Terminology Browser - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://nciterms.nci.nih.gov/NCIBrowser/Connect.do?bookmarktag=1&dictionary=NCI_Thesaurus&&code=C25164 Go

National Cancer Institute — U.S. National Institutes of Health | www.cancer.gov

Quick Search

Max Results 100

Advanced Search

Information about this concept

Accepted_Therapeutic_Use_For

Vocabulary: NCI_Thesaurus

Concepts visited (during this sessi

Date

QUICK LINKS

NCI HOME

NCICB HOME

EVS HOME

SITE HOME

KNOWN ISSUES

HELP | RESULTS | CUSTOMIZE | ABOUT | BROWSE HIERARCHY | LOGOUT

Concept Details

Bookmark this page

Date — Printable Page | History | Graph

Identifiers:

name	Date
code	C25164

Information about this concept:

DEFINITION	NCI|The particular day, month and year an event h happened or will happen.
Synonym with source data	DT|AB|CADSR
Synonym with source data	Date|PT|CADSR
Synonym with source data	Date|PT|NCI
Preferred_Name	Date
Semantic_Type	Temporal Concept
Synonym	DT
Synonym	Date
Unified Medical Language System Concept Identifier	C0011008

Trusted sites

caBIG™ cancer Biomedical Informatics Grid™ NATIONAL INSTITUTES OF HEALTH NATIONAL CANCER INSTITUTE

caARRAY

- strives to support existing standardization efforts and to follow caBIG guidelines to increase the probability that other software information systems can be interoperable with caARRAY

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caARRAY

- used MAGE UML model to describe caARRAY software system
- created MAGE UML class diagrams to illustrate classes, attributes, and relationships
- strived to use naming conventions and terminology standards prescribed by MAGE, MGED Society and the caBIG program
- strived to fully annotate MAGE UML model with class and attribute definitions, and with associated terminology concept codes

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caARRAY

- MAGE-OM API is well-documented
- MAGE-OM API provides information which is based upon an object-oriented abstraction of the underlying data in caARRAY
- MAGE-OM API returns data in the form of objects that are instances of classes in the UML model

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

caARRAY

- Mervi Heiskanen
- Anand Basu
- Xiaopeng Bian
- Jerry Eads
- Jim Zhou
- Hangjiong Chen
- Sumeet Muju
- Phu Tran
- Juergen Lorenz
- Scott Gustafson
- Himanso Sahni
- Durga Addepalli
- Ye Wu
- Don Swan
- Kevin Fitzpatrick
- John Moy
- Sharon Settnek
- Tara Akhavan

caBIG™ cancer Biomedical Informatics Grid™
NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

Thanks

- caBIG community