



caBIG[™] *cancer Biomedical
Informatics Grid*[™]

an initiative of the National Cancer Institute

LexGrid Vocabulary Services

Outline

- Overview
- Demonstration
 - LexBIG Services
- Related Work
 - caBIG™ Infrastructure
 - UK CancerGrid
 - NCBO BioPortal
- Q&A



caBIG™ cancer Biomedical
Informatics Grid™

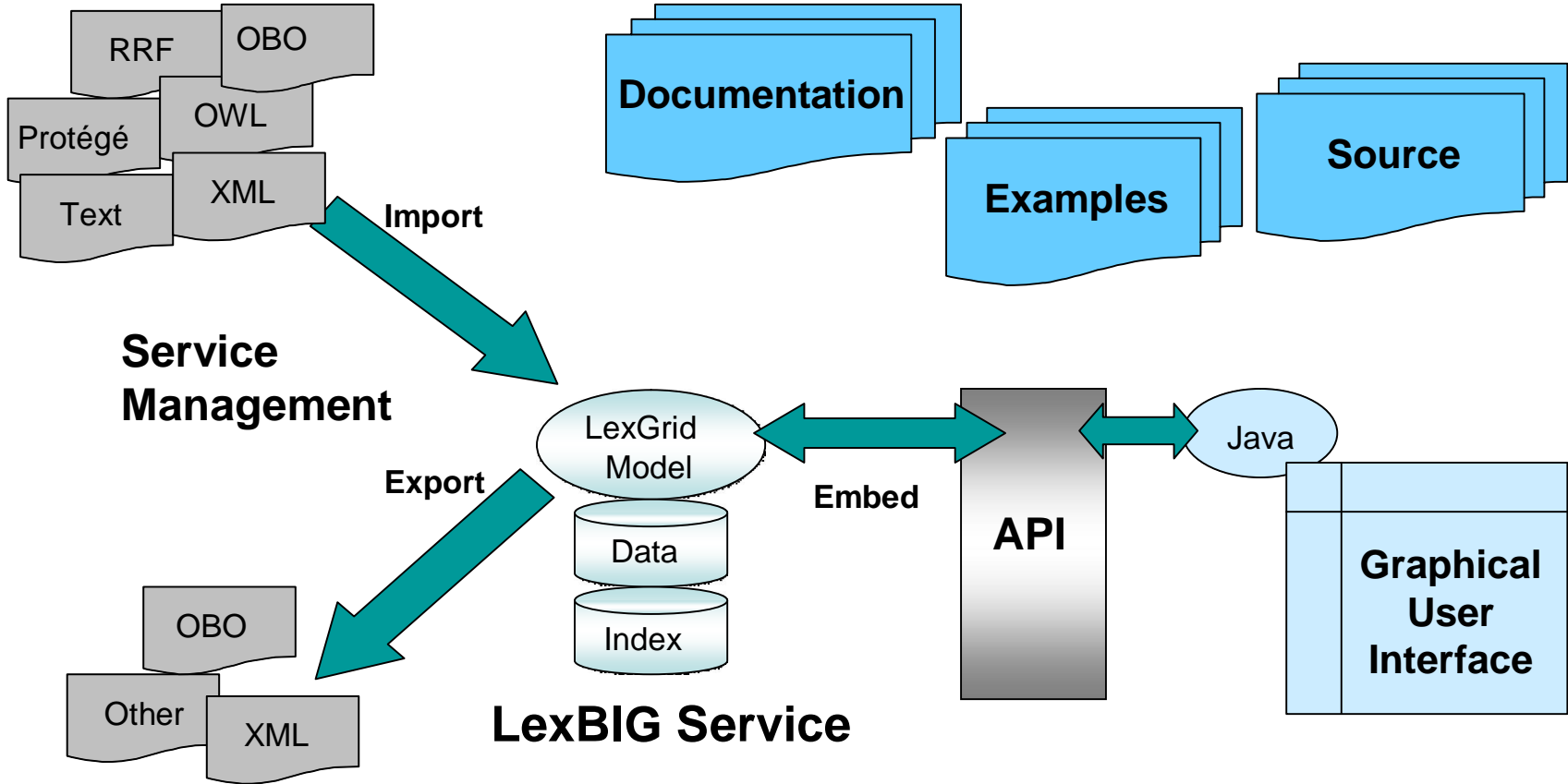
an initiative of the National Cancer Institute

Overview

- **LexGrid** represents an ongoing line of research ...
 - Provide information models and software components to represent and serve various vocabulary and ontology resources and formats (OWL, OBO, UMLS, Protégé frames, XML, proprietary, etc)
 - Grounded in standards (e.g. HL7 CTS, ISO 11179)
 - Not a single project or deliverable
- **HL7 Common Terminology Services (CTS) reference impl**
 - standards based effort
 - legacy
- **LexBIG-based projects**
 - **Moved away from CTS API (desire to align with CTS2)**
 - **LexBIG/LexBIO/LexPHIN are** specific software packages for vocabulary services, created to meet the requirements of the caBIG community, National Center for Biomedical Ontology, CDC, etc
 - Common open source code base



Overview



Overview

- LexBIG Packaging ...
 - **Service Management** tools to load, index, and manage vocabulary content for the vocabulary server.
 - **Application Programming Interface** (API) providing Java interfaces to various functions including Lexical queries, Graph representation and traversal, and NCI change event history.
 - **Graphical User Interface** providing access to service management and API functions; geared toward expert users.
 - **Documentation** consisting of JavaDocs, Administrator and Programmer Guides.
 - **Examples** providing sample source code for common vocabulary queries.
 - **Source** for all code developed for the LexBIG project.
 - **Test Suite** to validate the LexBIG installation.



Overview

- Not included in LexBIG ...
 - **Authoring or curation tools**
 - **End user tools** (friendly browser, etc)
 - **Application Server / Web Services**
 - **Vocabulary content**
 - **Data Repository**

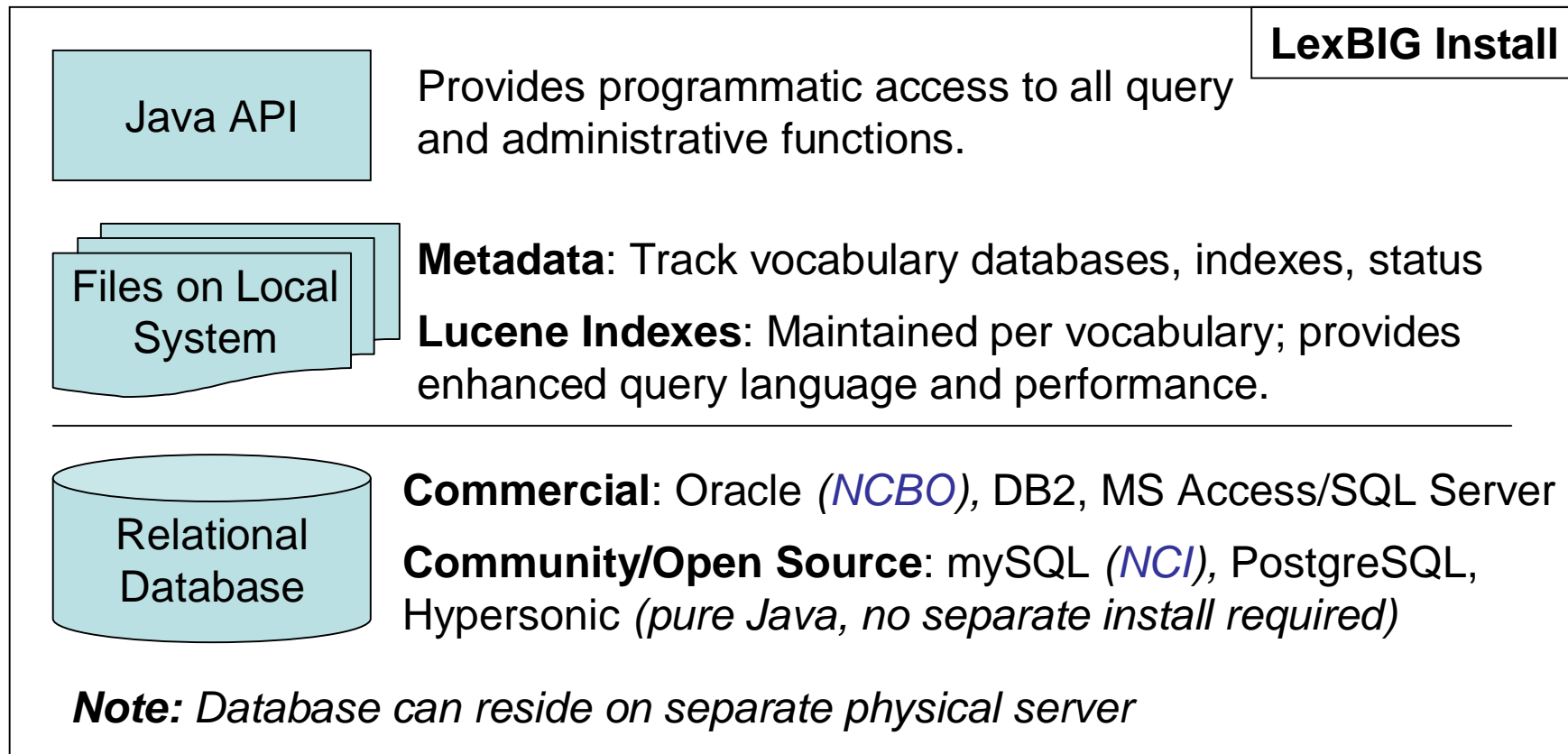


caBIG[™] *cancer Biomedical
Informatics Grid*[™]

an initiative of the National Cancer Institute

Configuration and Repository Access

- Each installation represents a LexGrid Node
- Each Node provides the software, metadata, indexes, and backing data store to service one or more vocabularies.



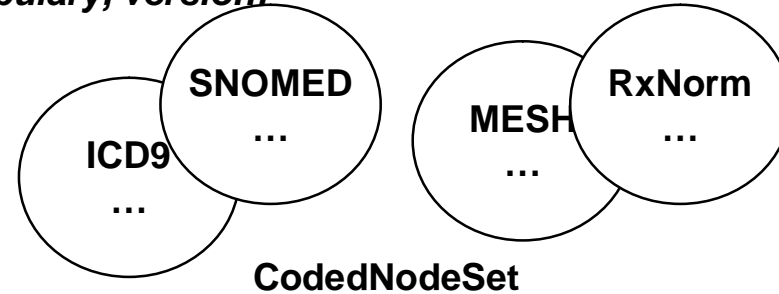
API Functionality

- Query Services
 - Search concepts (CodedNodeSet)
 - Search concept relationships (CodedNodeGraph)
 - Search concept history (NCI-specific)
 - Support for Iterators (on-demand resolution)
 - Support for match algorithms (e.g. starts-with, sounds-like)
 - Support for sorting, filtering
- Administrative Functions
 - Vocabulary import/export, and index maintenance
 - Vocabulary activation/deactivation, and tag assignment
- Extensible
 - Ability to register additional match algorithms, sort algorithms, filters, loaders, exporters, indexers, etc



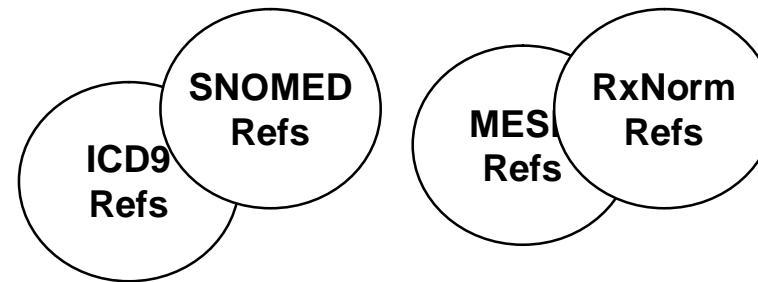
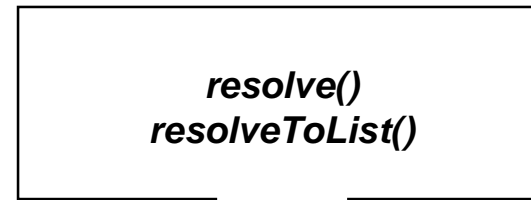
API Functionality – Query Concepts

`LexGridService.getCodingSchemeConcepts(vocabulary, version)`



`restrictToCodes()`
`restrictToCodeSystem()`
`restrictToMatchingProperties()`
`restrictToProperties()`
...

`union()`
`intersection()`
`difference()`



```

System.out.println("Example double restriction query with additional application of sort criteria and restricted
return values");

//Declare the service...
LexBIGService lbs = new LexBIGServiceImpl();

//Start with an unconstrained set of all codes for the vocabulary...
CodedNodeSet cns = lbs.getCodingSchemeConcepts("NCI_Thesaurus", null, false);

//Constrain to concepts with designations (assigned text presentations)
//that contain text that sounds like 'heart ventricle'
cns.restrictToMatchingDesignations("hart ventrikle",
SearchDesignationOption.ALL,
MatchAlgorithms.DoubleMetaphoneLuceneQuery.toString(),
null);

//Further restrict the results to concepts with a semantic type of
//'Anatomical Structure'.
cns.restrictToMatchingProperties(
Constructors.createLocalNameList("Semantic_Type"),
    "Anatomical Structure",
    "exactMatch",
null);

//Indicate that the resulting list should be sorted,
//with best results first and then sorted by code if there is a tie.
SortOptionList sortCriteria =
Constructors.createSortOptionList(new String[]{"matchToQuery", "code"});

//Indicate to return only the assigned UMLS_CUI and textualPresentation properties.
LocalNameList restrictTo =
ConvenienceMethods.createLocalNameList(new String[]{"UMLS_CUI",
    "textualPresentation"});

//Still nothing computed yet!
//Perform the query and resolve the sorted/filtered list,
//with a maximum of 6 items returned ...
ResolvedConceptReferenceList list = cns.resolveToList(sortCriteria, restrictTo, 6);


//Print the results ...
ResolvedConceptReference[] rcr = list.getResolvedConceptReference();
for(ResolvedConceptReference rc : rcr)
    System.out.println("Resolved Concept: " + ObjectToString.toString(rc));


```



Overview - Installation

Available on NCI GForge site ... <http://gforge.nci.nih.gov/projects/lexbig/>

National Cancer InstituteU.S. National Institutes of Health | www.cancer.gov

Center for Bioinformatics[Log Out](#) [My Account](#)


[Home](#) [My Page](#) [Project Tree](#) [Code Snippets](#) [Project Openings](#) [LexGrid Vocabulary Service](#)


[Summary](#) [Admin](#) [Forums](#) [Tracker](#) [Lists](#) [Tasks](#) [Docs](#) [News](#) [SCM](#) [Files](#)

Admin

Below is a list of all files of the project. Before downloading, you may want to read Release Notes and ChangeLog (access clicking on release version).

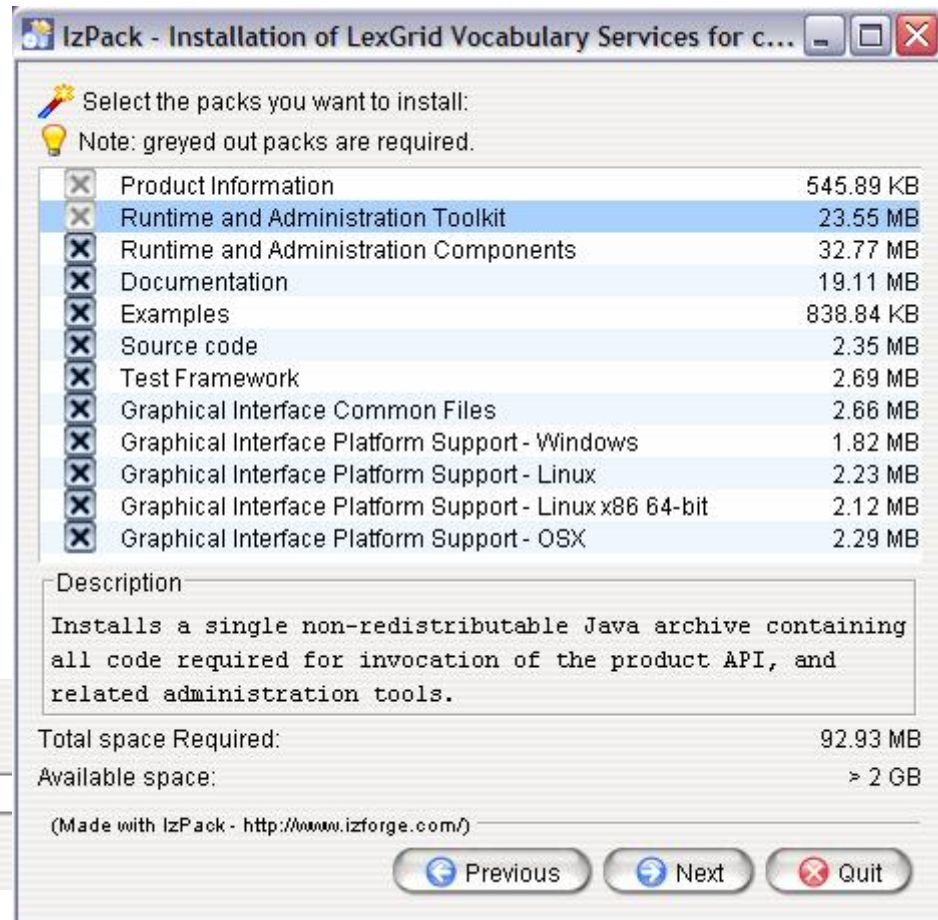
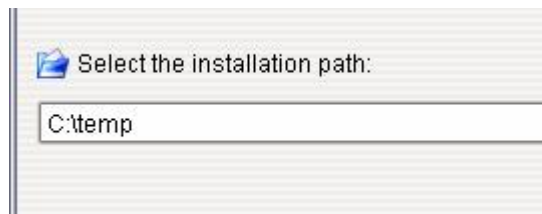
To create a new release click [here](#).

Package	Release & Notes	Filename	Size	Date	
				D/L	Arch
LexBIG 					
2.0Beta3					2007-01-18 15:35
	LexBIG-install-2.0Beta3.jar		72.83 MB	3 Any	.jar
	LexBIG-install-admin-guide-2.0Beta3.pdf		1.4 MB	4 Any	pdf
	LexBIG-install-config-2.0Beta3.xml		147 KB	3 Any	Other
	LexBIG-programmer-guide-2.0Beta3.pdf		1.06 MB	3 Any	pdf
1.0.1					2006-10-30 13:52

caBIG™ cancer Biomedical Informatics Grid™ an initiative of the National Cancer Institute

Overview - Installation

- Graphical Install

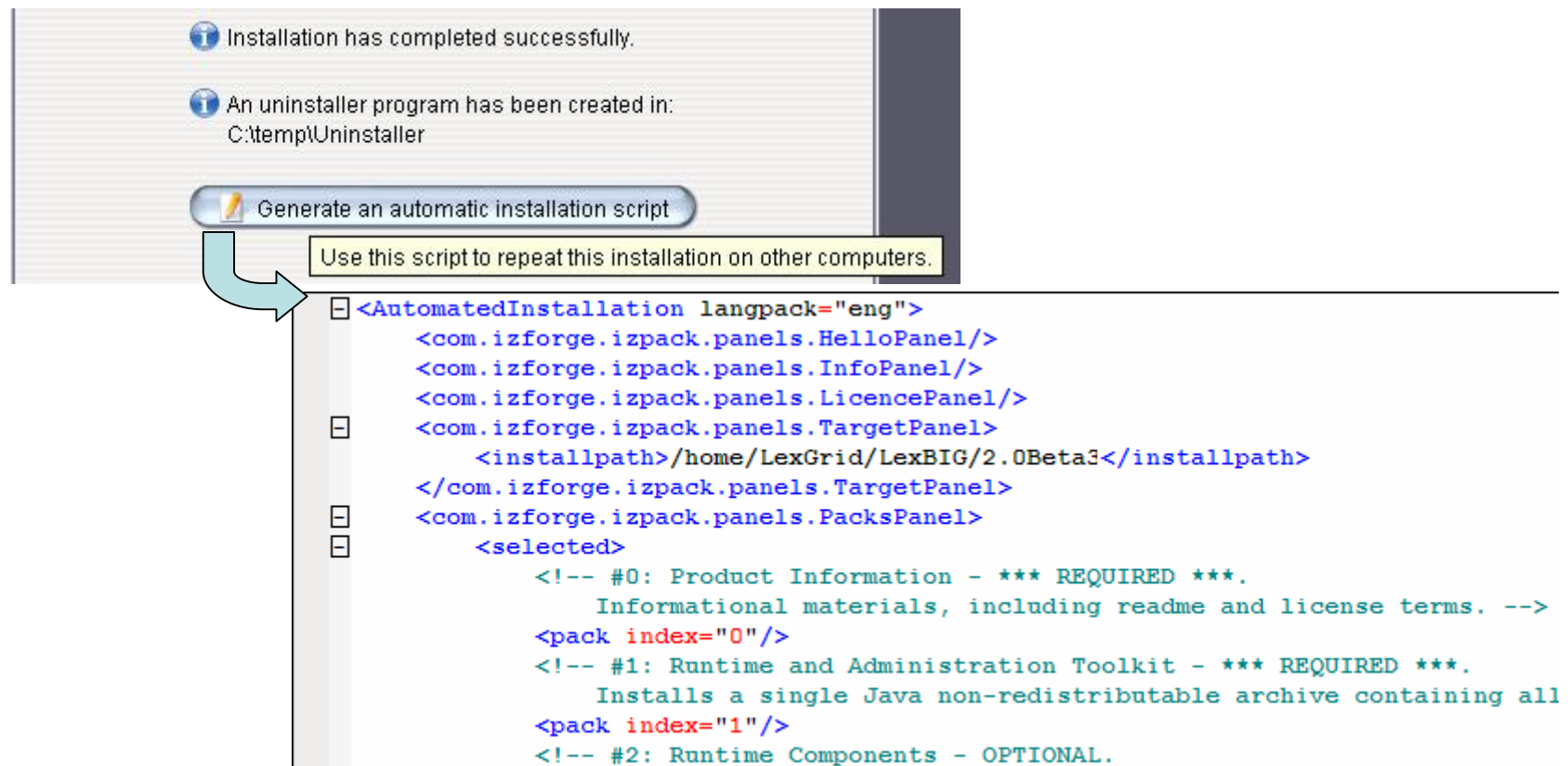


Overview - Installation

- Command Line Install

```
java -jar LexBig-install-xxx.jar installOptions.xml
```

- Simple to Automate



Installation has completed successfully.

An uninstaller program has been created in:
C:\temp\Uninstaller

Generate an automatic installation script

Use this script to repeat this installation on other computers.

```
<AutomatedInstallation langpack="eng">
  <com.izforge.izpack.panels.HelloPanel/>
  <com.izforge.izpack.panels.InfoPanel/>
  <com.izforge.izpack.panels.LicencePanel/>
  <com.izforge.izpack.panels.TargetPanel>
    <installpath>/home/LexGrid/LexBIG/2.0Beta3</installpath>
  </com.izforge.izpack.panels.TargetPanel>
  <com.izforge.izpack.panels.PacksPanel>
    <selected>
      <!-- #0: Product Information - *** REQUIRED ***.
        Informational materials, including readme and license terms. -->
      <pack index="0"/>
      <!-- #1: Runtime and Administration Toolkit - *** REQUIRED ***.
        Installs a single Java non-redistributable archive containing all
      <pack index="1"/>
      <!-- #2: Runtime Components - OPTIONAL.
```



API

<< Demonstration >>

Result Browser

Coding Scheme: UMLS_SemNet - urn:lsid.nlm.nih.gov:semnet
Concept Code: T001
Entity Description: Organism
Is Active: true
Presentation P0: Organism
Is Preferred: true

T001 - Organism
T028 - Gene or Genome
T026 - Cell Component
T190 - Anatomical Abnormality
T025 - Cell
T018 - Embryonic Structure
T024 - Tissue
T023 - Body Part, Organ, or Organ Component
T019 - Congenital Abnormality
T021 - Fully Formed Anatomical Structure
T020 - Acquired Abnormality
T017 - Anatomical Structure

has_part
has_part
has_part
has_part
has_part
has_part
has_part

T001
Organism

T024
Tissue

T023
Body Part, Organ, or Organ Component

T019
Congenital Abnormality



Testing and Serviceability

- Automated test suite

Unit Test Results

Designed for use with [JUnit](#) an

Class org.LexGrid.LexBIG.Impl.testUtility.AllTestsNormalConfig

Name	Tests	Errors	Failures	Time(s)
AllTestsNormalConfig	108	0	0	204.734

Tests

Name	Status	Type	Time(s)
testLoadAutomobiles	Success		44.297
testLoadGermanMadeParts	Success		25.516
testLoadHistory	Success		0.000
testLoadOwl	Success		26.562
testLoadGenericOwl	Success		13.703
testLoadNCIMeta	Success		20.047



caBIG[™] cancer Biomedical
Informatics Grid[™]

an initiative of the National Cancer Institute

Testing and Serviceability

- System Properties Report

basedir	C:\Program Files\LexGrid\LexBIG\2.0Beta3\test
DSTAMP	20070121
file.encoding	Cp1252
file.encoding.pkg	sun.io
file.separator	\
java.awt.graphicsenv	sun.awt.Win32GraphicsEnvironment
java.awt.printerjob	sun.awt.windows.WPrinterJob
java.class.path	.\lbTest.jar;..\runtime\lbPatch.jar;..\runtime\lbRuntime.jar;.\extlib\ant\ant.jar;.\extlib\ant\ant-launcher.jar;.\extlib\junit\junit.jar;C:\Program Files\LexGrid\LexBIG\2.0Beta3\test\extlib\ant\ant-launcher.jar;C:\Program Files\LexGrid\LexBIG\2.0Beta3\test\extlib\ant\ant-trax.jar;C:\Program Files\LexGrid\LexBIG\2.0Beta3\test\extlib\ant\ant-trax.jar
java.class.version	49.0
java.endorsed.dirs	C:\Program Files\Java\jre1.5.0_09\lib\endorsed
java.ext.dirs	C:\Program Files\Java\jre1.5.0_09\lib\ext
java.home	C:\Program Files\Java\jre1.5.0_09



Related Work

UK CancerGrid Consortium

<http://www.cancergrid.org/>



caBIG[™] cancer Biomedical
Informatics Grid[™]

an initiative of the National Cancer Institute

CancerGrid - Organization

CancerGrid is a consortium of leading researchers from the UK cancer and computing science communities developing open standards for clinical cancer informatics.

*LexBIG
Deployment*

Department of Oncology, University of Cambridge
Weatherall Institute of Molecular Medicine, University of Oxford
Division of Cancer Studies, University of Birmingham
Department of Oncology, University College London
Centre for Cancer Research and Cell Biology, Queen's University Belfast

Oxford University Computing Laboratory
Cambridge e-Science Centre
Midlands e-Science Centre
Belfast e-Science Centre



caBIG™ cancer Biomedical
Informatics Grid™

an initiative of the National Cancer Institute

Related Work

National Center for Biomedical Ontology

<http://www.bioontology.org/>



caBIG[™] cancer Biomedical
Informatics Grid[™]

an initiative of the National Cancer Institute

NCBO - Organization

Consortium of leading **biologists, clinicians, informaticians, and ontologists** who develop innovative technology and methods that allow scientists to create, disseminate, and manage biomedical information and knowledge in machine-processable form.

Vision - all biomedical knowledge and data are disseminated on the Internet using *principled ontologies*, such that the knowledge and data are *semantically interoperable* and useful for furthering biomedical science and clinical care.

Plans

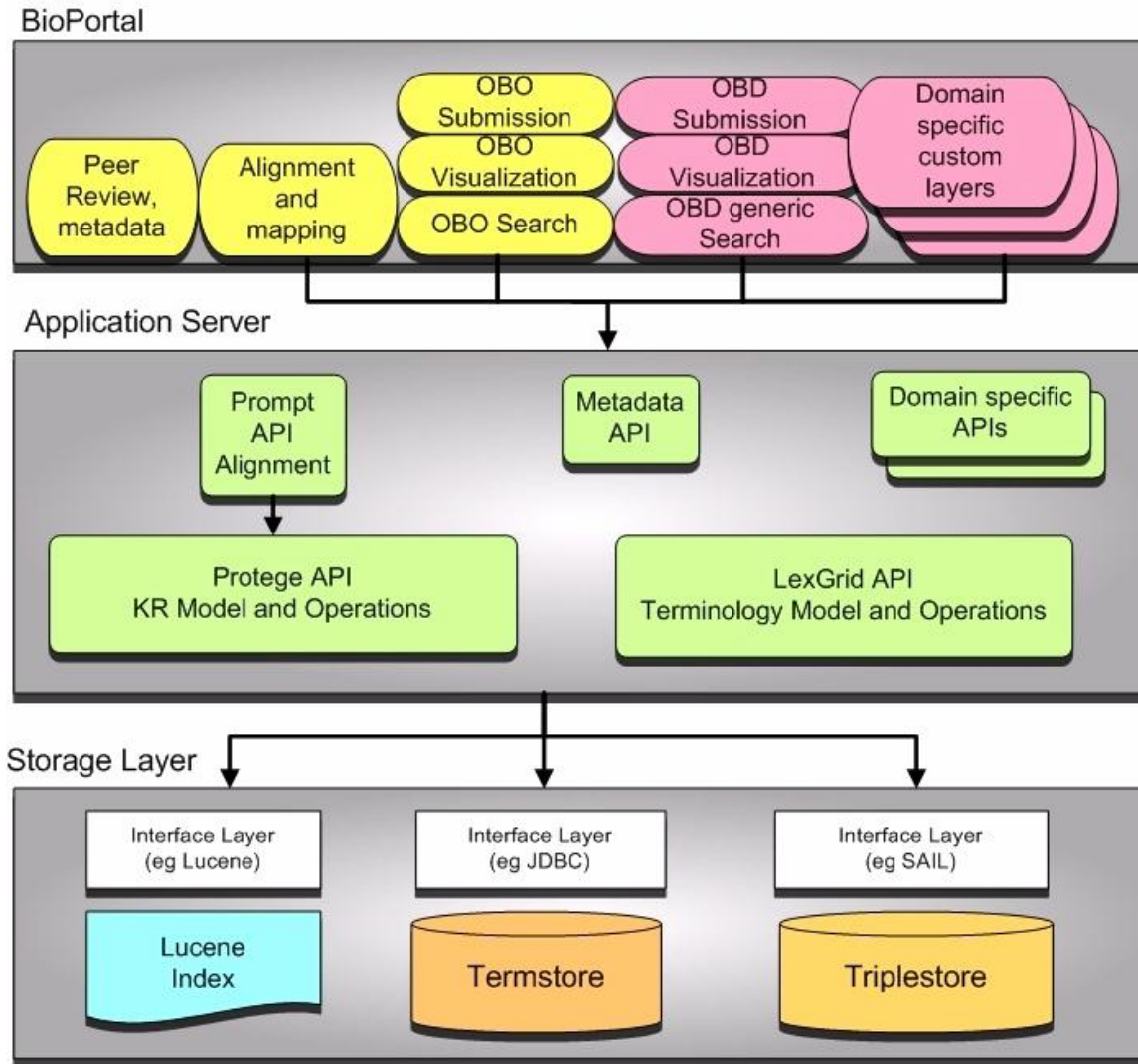
- URI-based access to ontology resources
- Web services to expose ontology resources
- Support ontology marginalia (glosses, annotations, diagrams)
- Support mapping and diff in context of ontology development/evolution



caBIG[™] cancer Biomedical
Informatics Grid[™]

an initiative of the National Cancer Institute

NCBO – BioPortal Architecture



NCBO – BioPortal (Demo)

The screenshot displays two AOL Explorer windows. The left window, titled 'Happy Page', shows a search interface for the NCBO BioPortal. The search criteria is 'gov.nih.nci.evs.domain.DescLogicConcept'. The right window, titled 'Result.jsp', displays the search results for the concept C12434. The results include a table of inverse role collections and a table of associated vocabularies.

Name	Value
Anatomic_Structure_is_Physical_Part_of	Peripheral_Blood_Cell
Anatomic_Structure_is_Physical_Part_of	Plasma
Gene_Product_Expressed_In_Tissue	DNA_Mismatch_Repair_Protein-MSH2

NamespaceId	Name	Description	SecurityToken								
40010	NCI_Thesaurus	NCI_Thesaurus	<table border="1"> <thead> <tr> <th>UserName</th> <th>Password</th> <th>AccessToken</th> <th>Certificate</th> </tr> </thead> <tbody> <tr> <td>null</td> <td>null</td> <td>null</td> <td>null</td> </tr> </tbody> </table>	UserName	Password	AccessToken	Certificate	null	null	null	null
UserName	Password	AccessToken	Certificate								
null	null	null	null								

IsA	TraverseDown	Name	Links
false	true	null	

Name	Value	Qualifi
FULL_SYN	peripheral blood	



Future Direction

- API extensions
 - Support for logical named groupings of concepts (value sets)
 - Support submission of vocabulary change requests
 - Support to load and map additional vocabulary formats
 - Support local vocabulary extensions
- API architecture
 - Provide a light-weight Representational State Transfer (REST) service implementation.
 - Synchronization and registration/discovery of services
 - Reasoner or classifier adaptation
- Alignment and advancement of standards (HL7/CTS2, ISO 11179, etc)
- Continued collaboration
 - caBIG™ – NCI, CancerGrid, NCBO, CDC/PHIN
- Provide framework for other parties to be more integrally involved with project direction and code development
 - **Standards & Specifications** – HL7
 - **Code** - Eclipse Open Healthcare Framework (OHF/OHIT)



Project Resources and Communication

- Mayo LexGrid Home Page:
<http://informatics.mayo.edu/LexGrid/>
- LexBIG GForge:
<http://gforge.nci.nih.gov/projects/lexbig/>
- EVS / LexBIG install:
<http://cabio-lexbig.nci.nih.gov/NCIBrowser/Dictionary.do>
<http://cabio-lexbig.nci.nih.gov/cacore32/Happy.jsp>
- EVS Home Page:
http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/vocabulary
- NCBO Home Page:
<http://www.biontology.org>
- UK CancerGrid Home Page:
<http://www.cancergrid.org>

