caBIG Compatibility Guidelines

August 25, 2004

ICR Face-to-Face Meeting

NATIONAL CANCER INSTITUTE

caBIG cancer Biomedical Informatics Grid

caBIG Compatibility Guidelines

- **What are the caBIG Compatibility Guidelines ?**
  - Provide Guidelines on caBIG Compatibility for:
    - caBIG participants
    - Interested stakeholders on caBIG compatibility
  - Set of definitions that caBIG participants can use to measure the maturity level of potential caBIG applications.

caBIG Compatibility Guidelines

- **How do they differ from caBIG Principles**
  - The caBIG Principles do not provide the guidelines for Applications to become caBIG compliant
  - The caBIG Principles are the core values of the program, and apply to projects which are funded by NCICB
  - Unfunded or vendor projects can be caBIG compliant (by following the Compatibility Guidelines) without following all the caBIG Principles
  - Both are important parts of a functioning, community-wide caBIG

caBIG Compatibility Guidelines

- **When will the Guidelines be complete?**
  - On-going effort
  - Evolving, “Living Document”
  - Driven by input and participation of the caBIG Community

caBIG Compatibility Guidelines

- **How are the Guidelines to be used?**
  - “Remediation” of legacy Applications
  - “Evaluation” of existing Applications
  - “Creation” of new Applications

caBIG Compatibility Guidelines

- **What is the structure of the Guidelines?**
  - Levels of Maturity:
    - **"Legacy"**
    - **"Bronze"**
    - **"Silver"**
    - **"Gold"**

caBIG Compatibility Guidelines

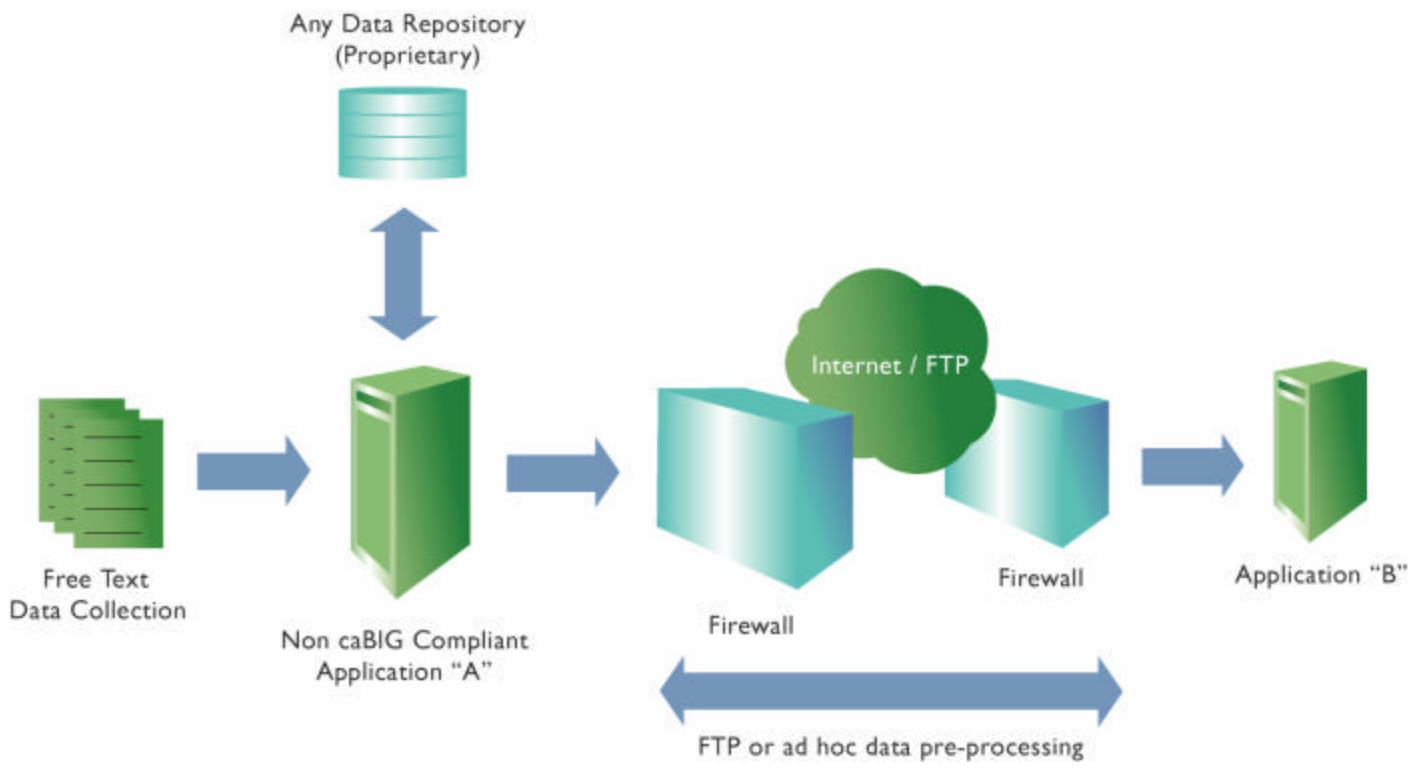
Grading - Matrix

Maturity Model	LEGACY	BRONZE	SILVER	GOLD
Interface Integration	-No Programming interfaces to the system are available. Only local data files in a custom format can be read -- ***Some ad hoc data transfer mechanism such as FTP***	***- Provide baseline* programmatic access to data. Data can be read from remote electronic sources or from commonly used file formats Data can be pushed out to from applications to other external data sources.***	-- Well-described API's that provide access to data objects. -- System architecture separated into tiers and interoperable components -- Data read in from standards-based electronic sources that support standard or commonly used interchange formats -- Documented component description of the underlying data structures that are accessible ***-- Standard messaging systems where appropriate***	- All features of Silver, plus: - Interoperable with data grid architecture to be defined by caBIG ***- Fully componentized provide access to individual resources in the form of grid services***
Vocabularies / Terminologies & Ontologies	***- Free text used throughout for data collection***	***- Use of publicly accessible standardized controlled vocabularies as well as local terminologies***	***- Messages need to be based on a common messaging model (such as HL7)***	- All features of Silver, plus: - Fully compliant with caBIG recommended standards for vocabulary terminology services and content sources ***- Common Data Elements should be built using controlled terminology***
Data Elements	***- No Structured metadata is recorded***	***- Some type of metadata describing the information in the system is used for data collection and external reporting. Metadata is retrieved from external repository shared by multiple applications.***	- Use common standard electronic representation for CDE's such as ISO 11179 or comparable standard ***- CDEs are harmonized and re-used from across the Domain Workspace***	- All features of Silver, plus: - Programmatic access to all metadata, including data class descriptions, site and source information, and any other caBIG-defined metadata requirements and use information models ***- Use the caBIG standard or electronic representation of metadata and Common Data Elements***
Information Models	***- No particular information model is used to represent data***	***- Some type of diagrammatic model describing the data relationship is available in electronic format***	***- Information models defined in a standard modeling language such as UML***	- All features of Silver, plus: ***- Information models are harmonized with other s across the caBIG Domain Workspace***

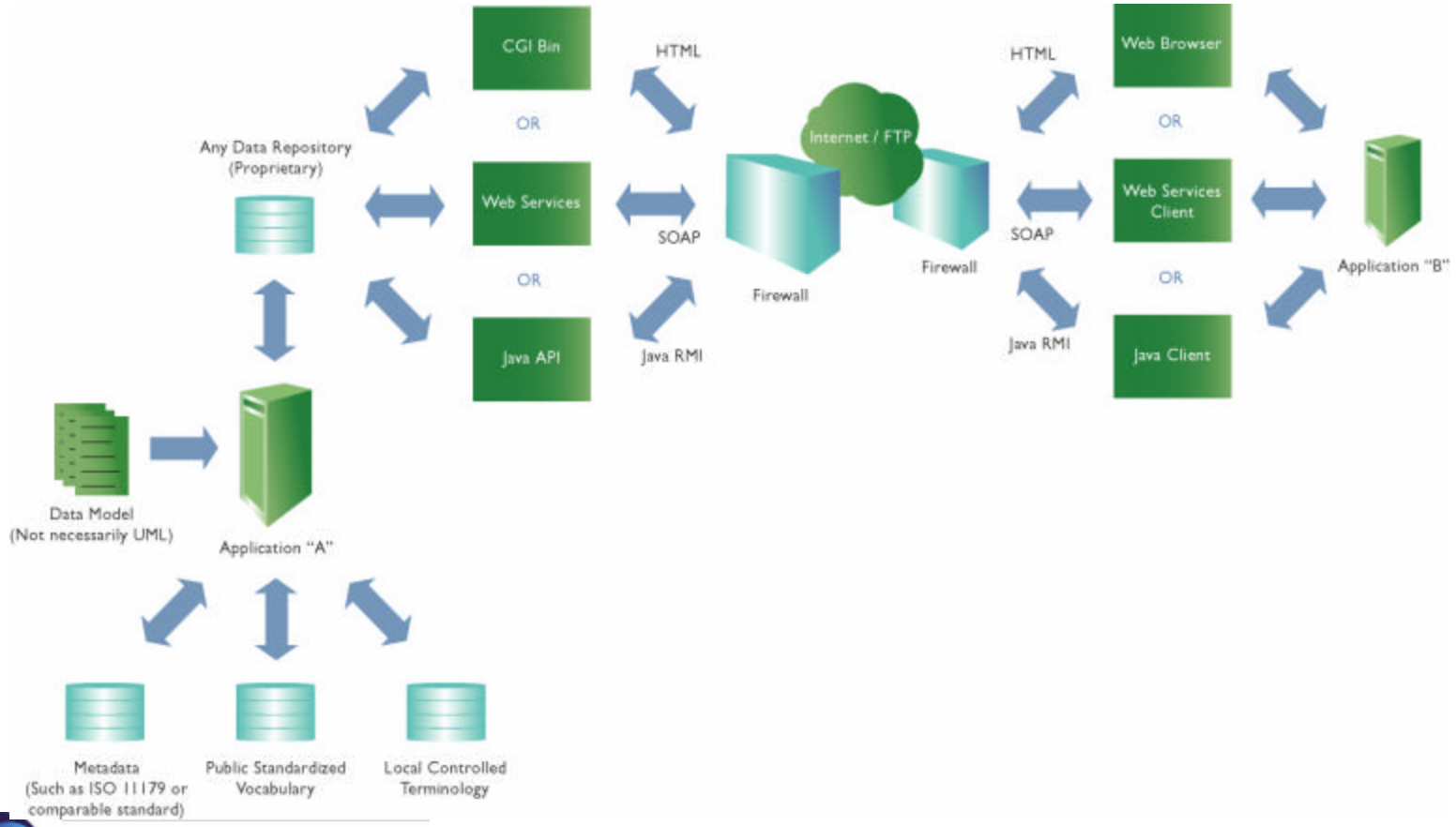
caBIG Compatibility Guidelines

- **Requirements**
  - Minimal requirements – ”Bronze”
  - New Development **“Silver”**
  - Deprecating ”Bronze” in c. 24 Months
  - caBIG Community defines **“Gold”**

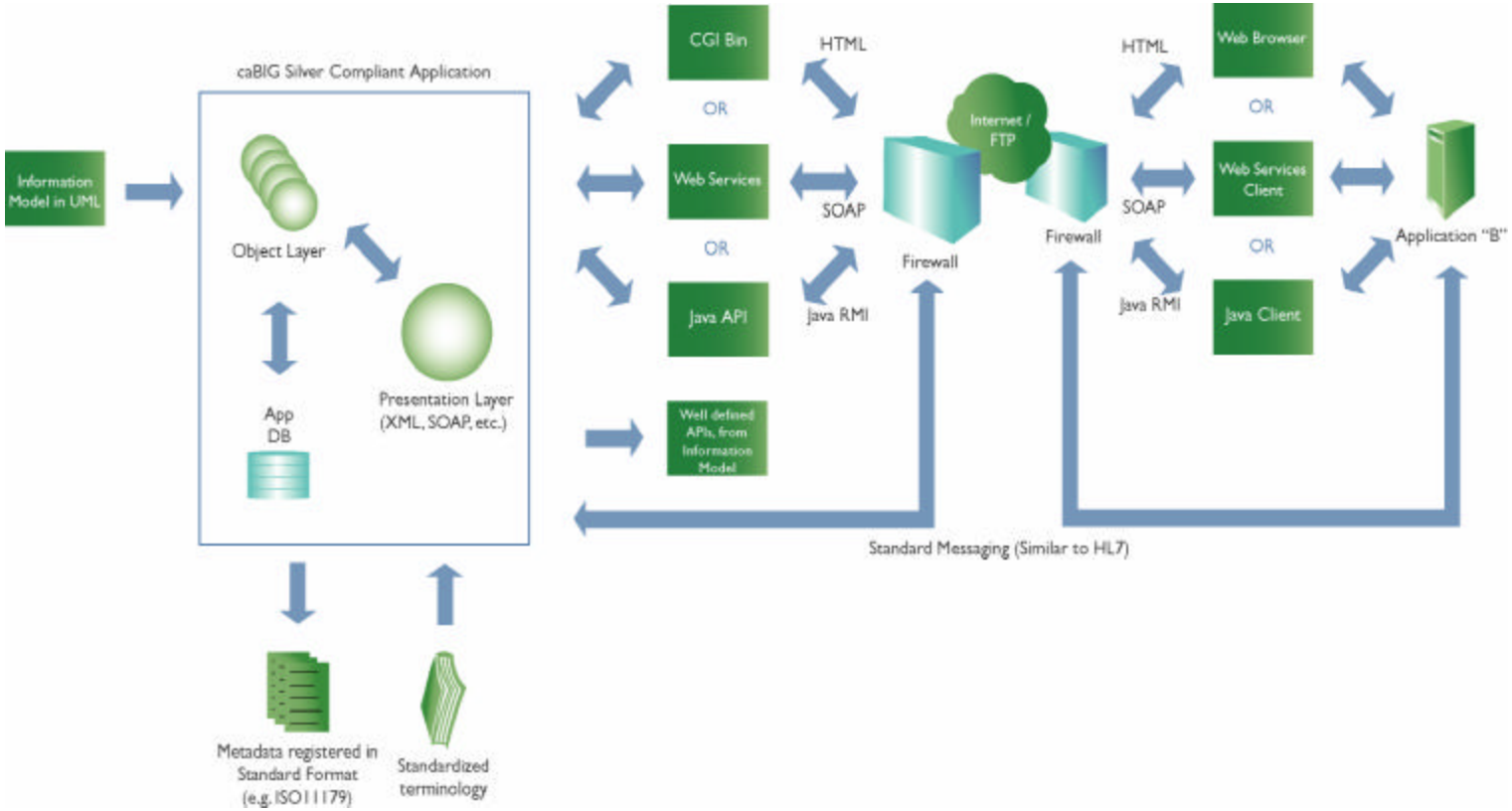
Outline of “Legacy” Level of Maturity

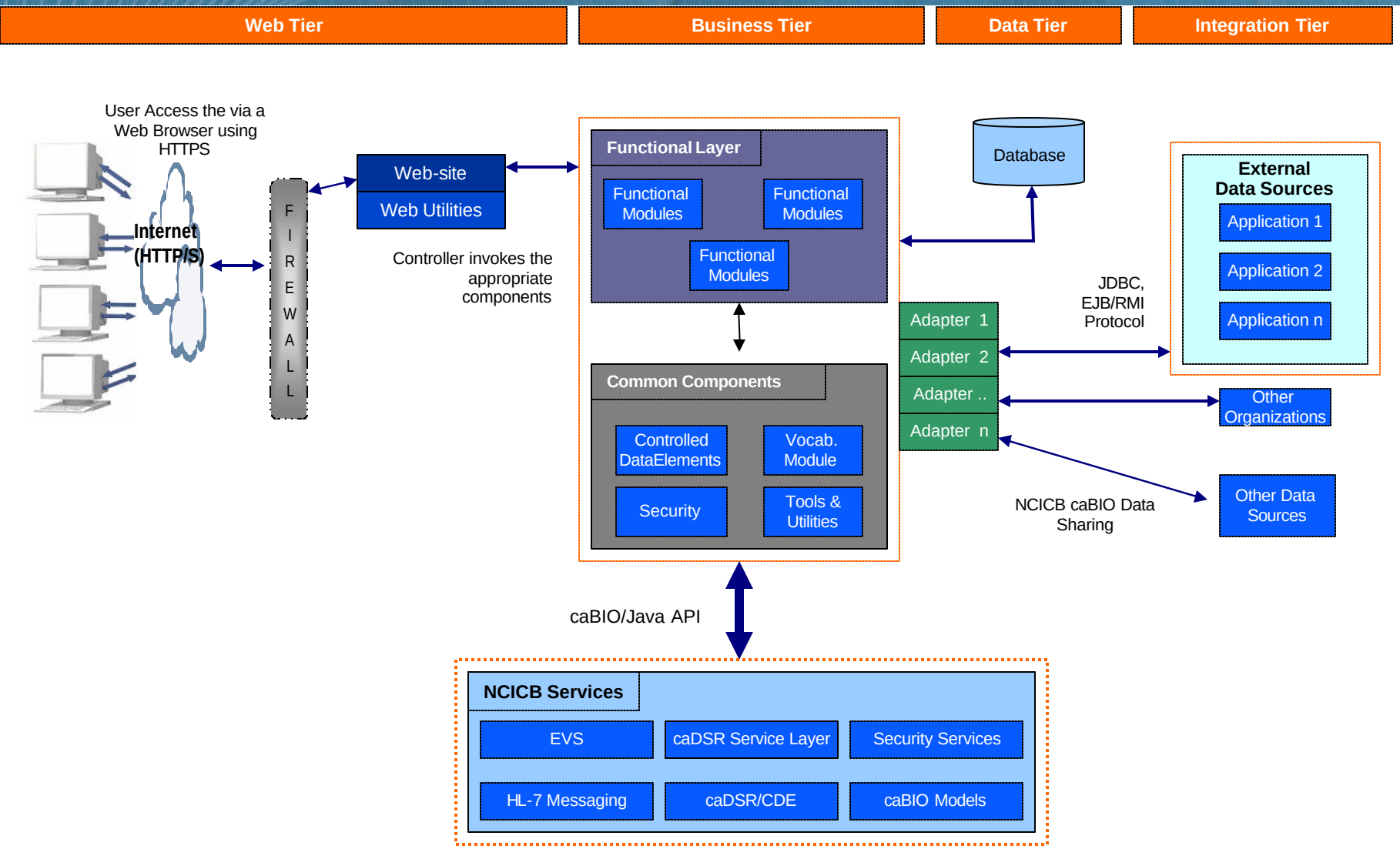


Outline of "Bronze" Level of Maturity



Outline of "Silver" Level of Maturity





NATIONAL CANCER INSTITUTE

caBIG cancer Biomedical Informatics Grid

BRONZE MATURITY LEVEL

Interface Integration

Access to data electronically by using industry standard file formats.

Examples:

JDBC – Access to external database using JDBC over TCP/IP

Web Services – Invoke application level API using SOAP

Vocabularies/Terminologies:

Usage of Public standardized controlled vocabularies.

Examples:

Manual input of Vocabulary terms when submitting or editing data

Data Elements:

Metadata describing information in the system, stored in electronic format.

Examples:

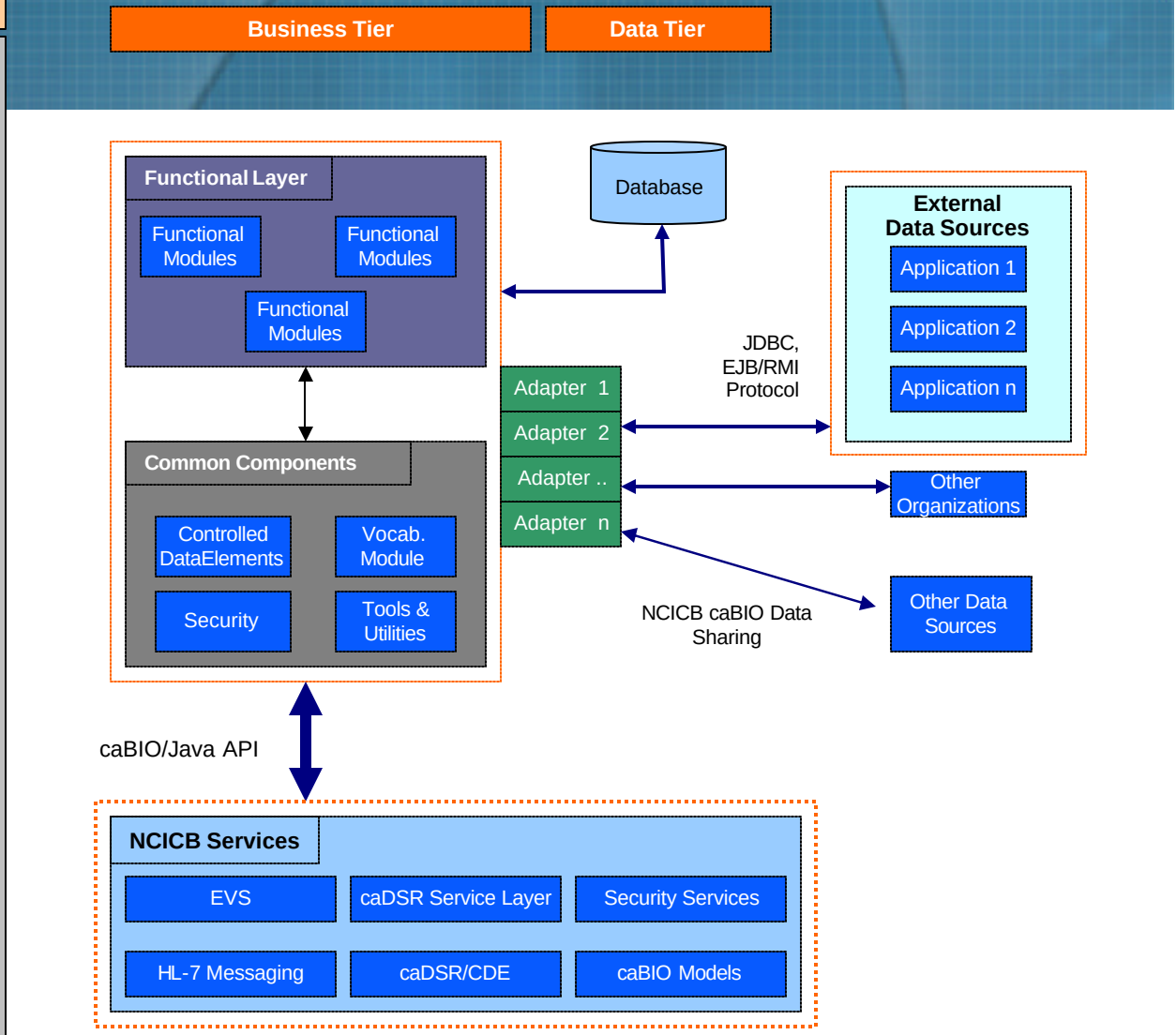
Controlled spreadsheet containing Data Elements used in the system

Information Model:

Diagrammatic model describing the data relationship available in electronic format.

Examples:

Entity Relationship Diagram of the data model



BRONZE LEVEL - INTERFACE INTEGRATION

Access to data electronically by using industry standard file formats.

Examples: JDBC – Access to external database using JDBC over TCP/IP

JDBC Code Sample

*CSAERS DAO Code Sample

```
    // Initialize Remote Connection using JDBC 2.0 and Prepared Statement
    // create the prepared statement
    try {
        prepStmt = conn.prepareStatement(userPersonalNoticeQueryString);
        // fill in the parameters from the query
        prepStmt.clearParameters();
        prepStmt.setInt(1, userVO.getMessageId().intValue());
        prepStmt.setString(2, userVO.getUserId());
        prepStmt.setString(3, userVO.getMessageText());
        prepStmt.setString(4, userVO.getCreatedBy());
        prepStmt.setTimestamp(5, new Timestamp(userVO.getCreatedDate().getTime()));
        prepStmt.setTimestamp(6, new Timestamp(userVO.getExpirationDate().getTime()));
        prepStmt.setInt(7, userVO.getMessageDefinitionId().intValue());
        prepStmt.setInt(8, userVO.getCaseId().intValue());

        // execute the stmt
        resultSet = prepStmt.executeQuery();

    } catch (SQLException sql_e) {
        String message = "Cannot execute personal notices list query";
        logger.error(message, sql_e);
        throw new DAOAccessException(message, sql_e);
    } finally {
        // Close Result Set, Statement and Connection
    }
}
```

BRONZE LEVEL - INTERFACE INTEGRATION

Access to data electronically by using industry standard file formats.

Example: Web Services – Invoke application level API using SOAP e.g., caBIO SOAP API

Web Services Code Sample

*caBIO SOAP Example, using SOAP::LITE

```
$URI='urn:nci-gene-service';
$PROXY_PATH='/soap/servlet/rpcrouter';
my $argc=@ARGV;
# anonymous hash reference, passed as map to SOAP service
my $searchRec={};
# first three arguments are mandatory
if ($argc < 3) {
    die $USAGE;
} else {
    $server = shift @ARGV;
    $port = shift @ARGV;
    $method = shift @ARGV;
    $argc=@ARGV;
    while ($argc) {
        my $arg = shift @ARGV;
        if ($arg eq "-symbol") {$searchRec->{symbol} = shift @ARGV;}
        elsif ($arg eq "-genBankAccessionNumber") {$searchRec->{genBankAccessionNumber} = shift @ARGV;}
        elsif ($arg eq "-pathwayId") {$searchRec->{pathwayId} = shift @ARGV;}
        ...
        ...
        elsif ($arg eq "-returnHeavyXML") {$searchRec->{returnHeavyXML} = shift @ARGV;}
        elsif ($arg eq "-fillInObjects") {$searchRec->{fillInObjects} = shift @ARGV;}
        elsif ($arg eq "-targetId") {$searchRec->{targetId} = shift @ARGV;}
       else {
         print "\n****Please check the name of attribute****\n";
         die $USAGE;
        }
        $argc=@ARGV;
    }
}

$s = SOAP::Lite -> uri($URI) -> proxy("http://$server:$port$PROXY_PATH");

# make service request
$som=$s->$method(SOAP::Data->type(map => $searchRec));
# interpret result
if ($som->fault) {
    print "FAULT ENCOUNTERED!\nfaultcode:\t" . $som->faultcode . "\nfaultstring:\t" . $som->faultstring . "\n";
} else {
    $xmldoc = $som->result;
    print "\nMETHOD CALLED: $method\n\n" . decode_entities($xmldoc) . "\n\n";
}
```

BRONZE LEVEL -VOCABULARIES

Usage of Public standardized controlled vocabularies.

Example: EVS Lookup or Manual look-up using internal Medical Coding Libraries

EVS Code Sample

*caBIO EVS Code Samples

```
DescLogicConceptSearchCriteria dlcsc = new DescLogicConceptSearchCriteria();
DescLogicConcept dlc = new DescLogicConcept();
MetaThesaurusConceptSearchCriteria mtcsc = new MetaThesaurusConceptSearchCriteria();
MetaThesaurusConcept mtc = new MetaThesaurusConcept();
String vocabularyName= "NCI_Thesaurus";
String conceptname = "Gene";

//-------------------------------------------------------------------------
//     NCI Thesaurus
//     search()
//-------------------------------------------------------------------------

dlcsc.setLimit(10); //default is 100
dlcsc.setSearchTerm("Gen*");
dlcsc.setVocabularyName("NCI_Thesaurus");
Concept[] conceptArray = dlc.search(dlcsc);
  if(conceptArray!=null) {
    for(int i=0; i<conceptArray.length; i++) {
        System.out.println(conceptArray[i].getName());
    }
  }
```

BRONZE LEVEL –DATA ELEMENTS

Metadata describing information in the system, stored in electronic format

Examples: Controlled spreadsheet containing Data Elements used in the system

Data Element Sample

Element Name	Element Type	Minimum Length	Maximum Length	Associated Page	Data Format Validation	Required	Allowed Values
Patient_First_Name	String	1	100	REGISTRATION	N/A	✓	N/A
Patient_Last_Name	String	1	100	REGISTRATION		✓	N/A
Patient_Suffix	String	0	5	REGISTRATION		✓	Mr. Mrs. Ms.
Patient_Address_Line_1	String	1	200	REGISTRATION		✓	
Patient_Address_Line_2	String	1	200	REGISTRATION		✓	

Controlled Excel Spreadsheet to manage Data Elements associated with the System

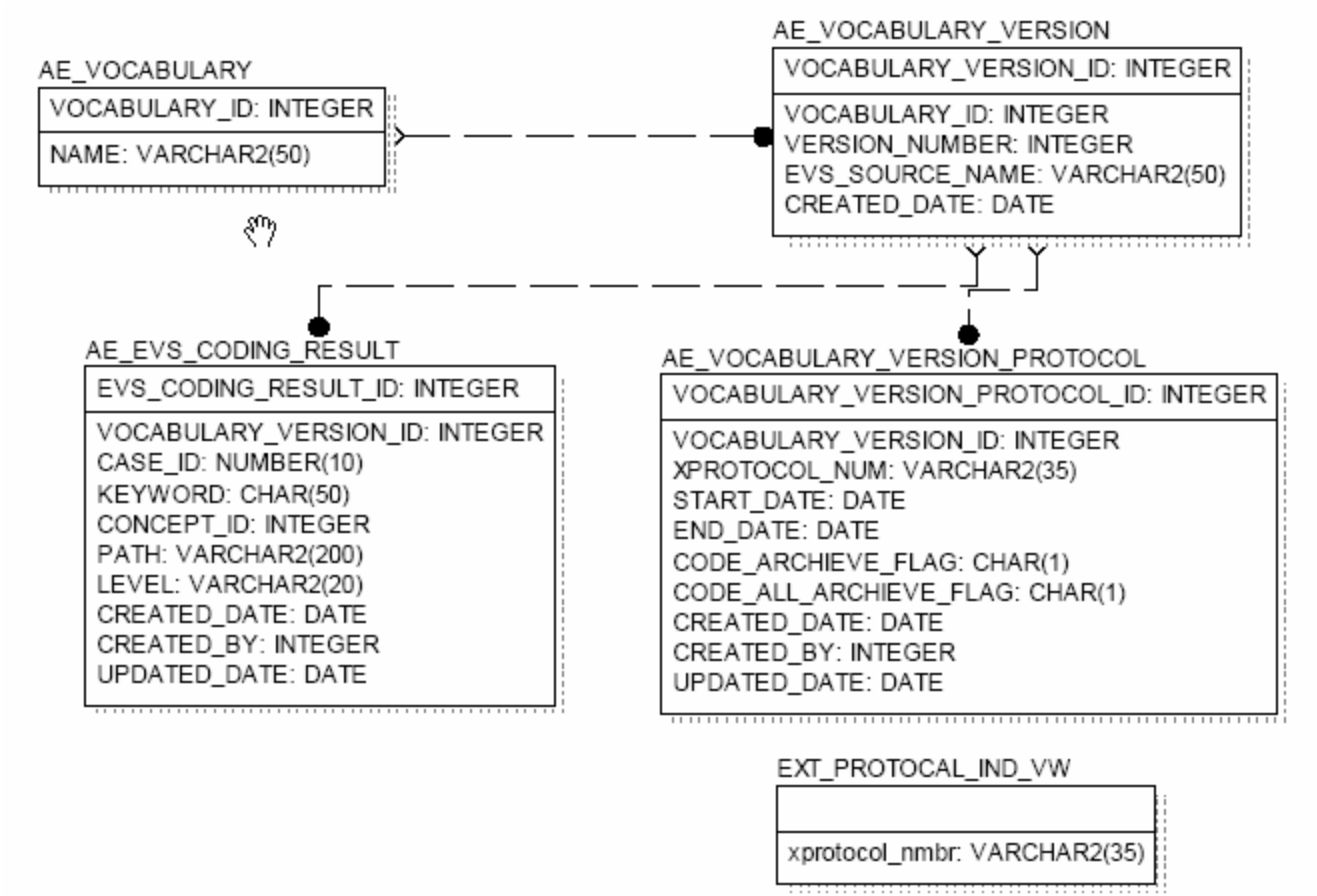
BRONZE LEVEL – INFORMATION MODEL

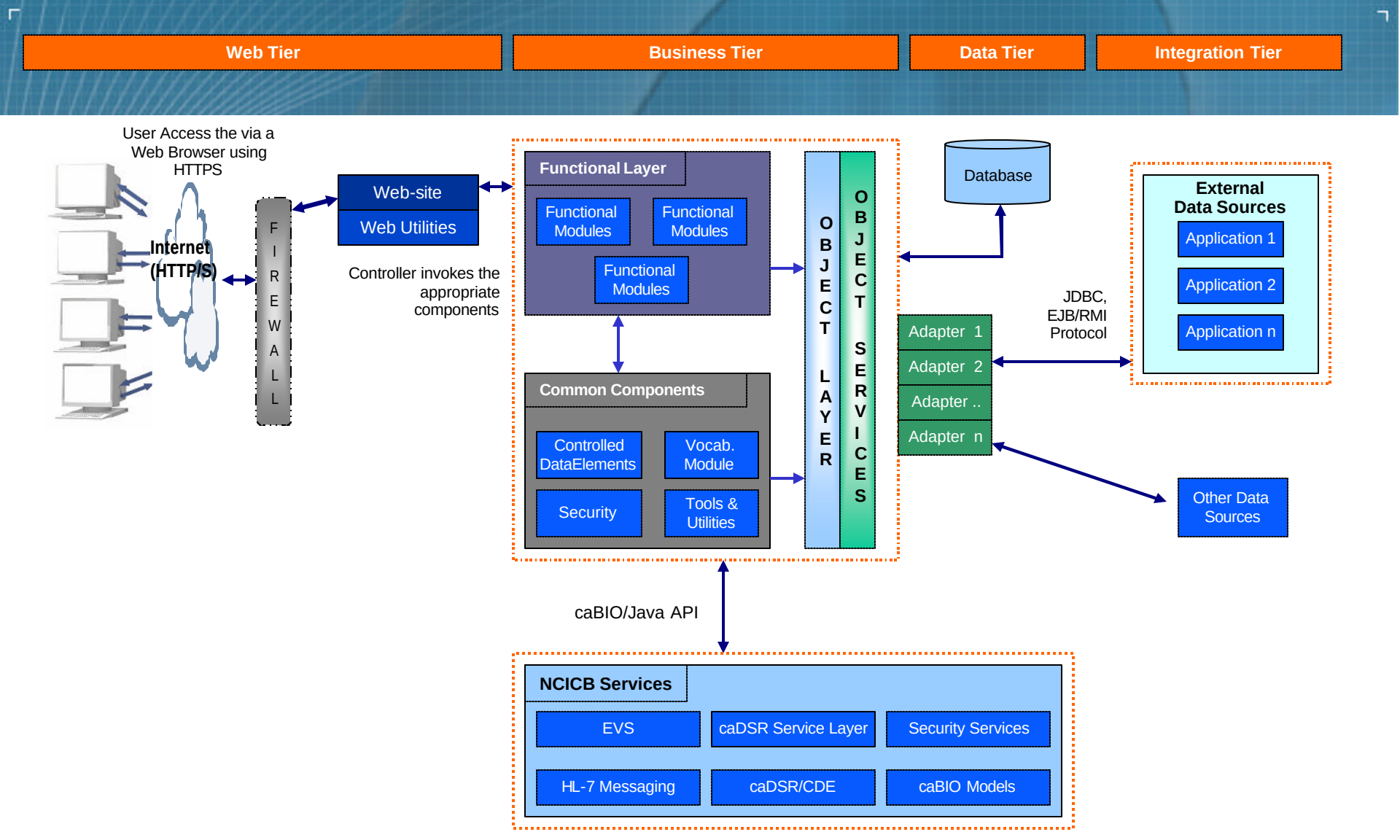
Diagrammatic model describing the data relationship available in electronic format.

Examples: Entity Relationship Diagram of the data model

Data Element Sample

*CSAERS EVS Medical Coding ERWin Diagram




Web Tier
Business Tier
Data Tier
Integration Tier
User Access the via a Web Browser using HTTPS
Internet (HTTP/S)
FIREWALL
Web-site
Web Utilities
Controller invokes the appropriate components
Functional Layer
Functional Modules
Functional Modules
Functional Modules
Common Components
Controlled DataElements
Vocab. Module
Security
Tools & Utilities
OBJECT LAYER
OBJECT SERVICES
Database
External Data Sources
Application 1
Application 2
Application n
JDBC, EJB/RMI Protocol
Adapter 1
Adapter 2
Adapter ..
Adapter n
Other Data Sources
caBIO/Java API
NCICB Services
EVS
caDSR Service Layer
Security Services
HL-7 Messaging
caDSR/CDE
caBIO Models

SILVER MATURITY LEVEL

Interface Integration

Well described API to provide access to data objects

System architecture separated into tiers

Data read from standard electronic sources that support standard or commonly used interchange formats

Documented component description of the underlying data structures

Standard messaging systems where appropriate

Vocabularies/Terminologies:

Usage of standard terminologies approved by public standard bodies or caBIG Vocabulary/CDE Workspace

Data Elements:

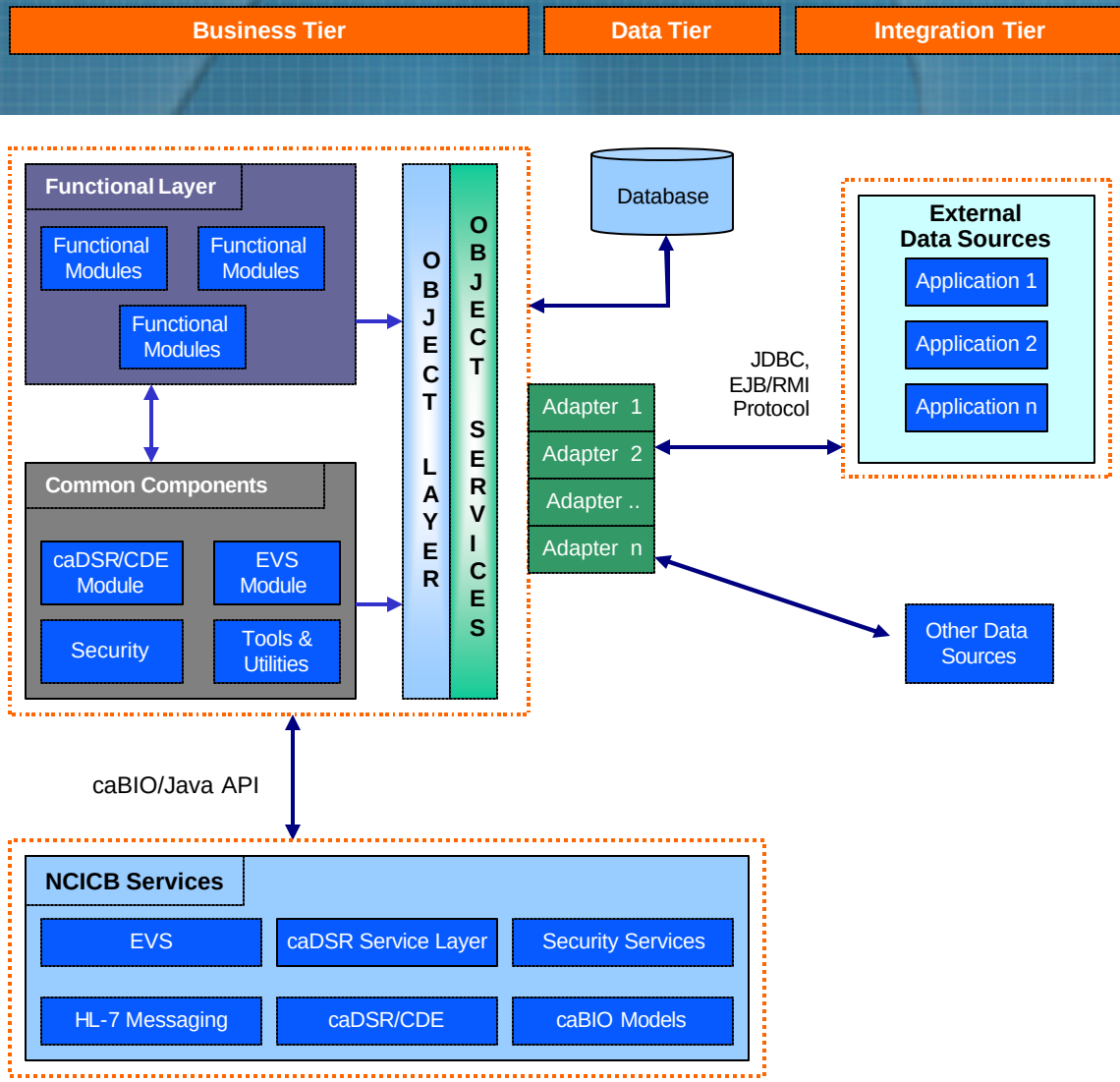
Use common standard electronic representation for CDE's such as ISO 11179 or comparable standard

Harmonization and re-use of CDEs

CDEs built using standard controlled terminologies approved by caBIG Vocabulary/CDE Workspace

Information Model:

Use of standard modeling language such as UML to represent the information model



SILVER LEVEL - INTERFACE INTEGRATION

Well described API to provide access to data objects

Examples: caBIO API accessing Gene data object

Data Objects API Code Sample

*caBIO Code Sample

```
Gene myGene = new Gene();
GeneSearchCriteria criteria = new GeneSearchCriteria();
criteria.setSymbol("FECH");
SearchResult result = myGene.search(criteria);

if (result != null){
  Gene[] genes = (Gene[]) result.getResultSet();
    for(int i = 0; i < genes.length; i++){
      System.out.println("\nInformation regarding Gene "+genes[i].getName());
      //System.out.println("\nId: "+genes[i].getId());
      System.out.println("\tTitle: "+genes[i].getTitle());
      System.out.println("\tOMIM Id: "+genes[i].getOMIMId());
      System.out.println("\tUnigene Cluster Id: "+genes[i].getClusterId());
      System.out.println("\tLocusLink Id: "+genes[i].getLocusLinkId());
      System.out.println("\tOrganism Abbreviation: "+genes[i].getOrganismAbbreviation());

      Sequence[] refSeqs = genes[i].getReferenceSequences();
      if( refSeqs.length > 0 ){
        System.out.println("\nAssociated Reference Sequence(s): ");
        for(int k = 0; k < refSeqs.length; k++){
          System.out.println("\n\tSequence Id:"+refSeqs[k].getId());
          System.out.println("\tSequence Accession Number:"+refSeqs[k].getAccessionNumber());
          System.out.println("\tSequence Type:"+refSeqs[k].getType());
          System.out.println("\tSequence Ascii String:"+refSeqs[k].getAsciiString());

          }
        } ...
```

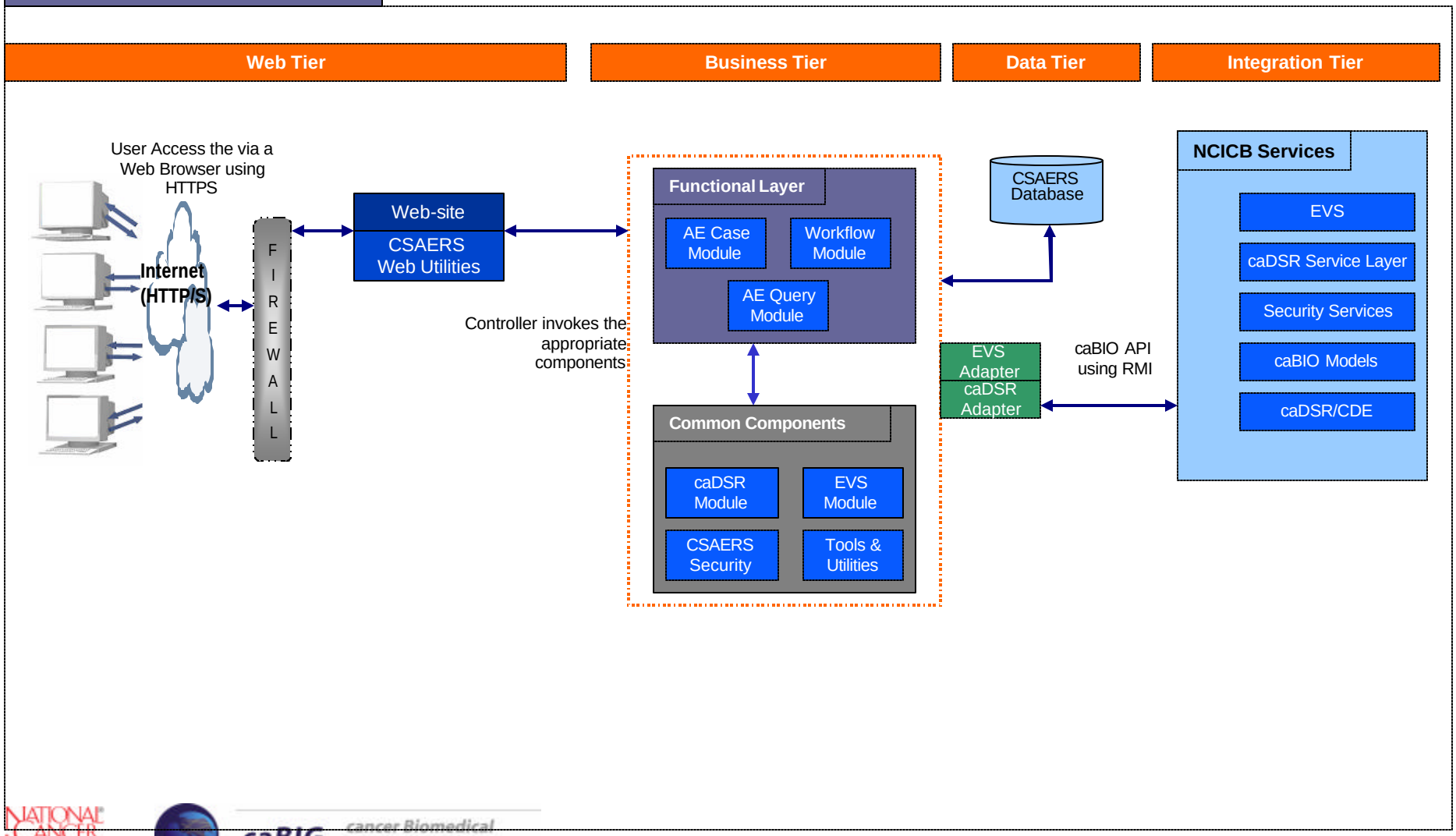
SILVER LEVEL - INTERFACE INTEGRATION

System architecture separated into tiers

Examples: CSAERS n-Tier Architecture Diagram

n-Tier Architecture Sample

*CSAERS n-Tier Architecture



SILVER LEVEL - INTERFACE INTEGRATION

Data read from standard electronic sources that support standard or commonly used interchange formats

Examples: Standard HL-7 electronic format used to read-in AE Events messages from multiple applications. HL-7 Ant script to load and run DemoDriver using the HL-7 v3 RIM Java SIG Implementation (alpha version)

Data Objects API Code Sample

```
<!– Ant Script to run HL-7 Java SIG Demo -->
<!– Demo Driver will use the RIM hmd.file.adverseEvent files -->
<!-- as the “Model” file and the msg.file.adverseEvent file as the message to transform -->

<java fork="yes" classname="org.hl7.demo.DemoDriver"
   failonerror="true"
   classpathref="project.class.path">
   <sysproperty key="org.xml.sax.driver"
              value="com.icl.saxon.aelfred.SAXDriver" />
   <jvmarg value="-Xbootclasspath/p:${lib.dir}/gjc-rt.jar"/>
   <sysproperty key="java.util.logging.config.file"
     value="logging.properties" />
   <arg value="-x"/>
   <arg value="${hmd.dir}/${hmd.file.adverseEvent}"/>
   <arg value="${rim.file}"/>
   <arg value="-f"/>
   <arg value="${etc.dir}/${msg.file.adverseEvent}"/>
   <arg value="-m"/>
   <arg value="${msg.type.adverseEvent}"/>
 </java>
</target>
```

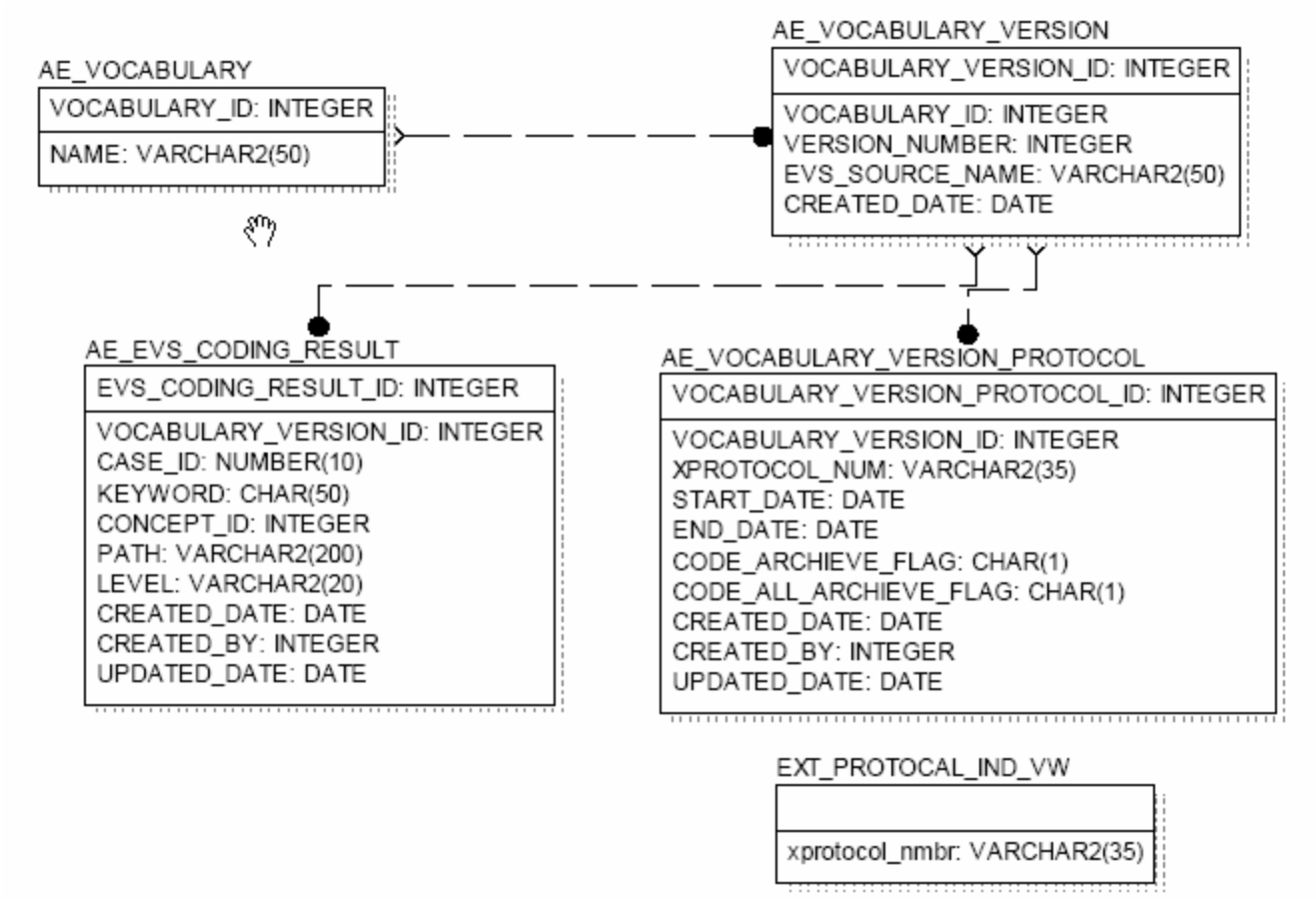
SILVER LEVEL - INTERFACE INTEGRATION

Documented component description of the underlying data structures

Examples: CSAERS Object Model for Medical Coding Module

Underlying Data Structure Example

*CSAERS Medical Coding Data Structure

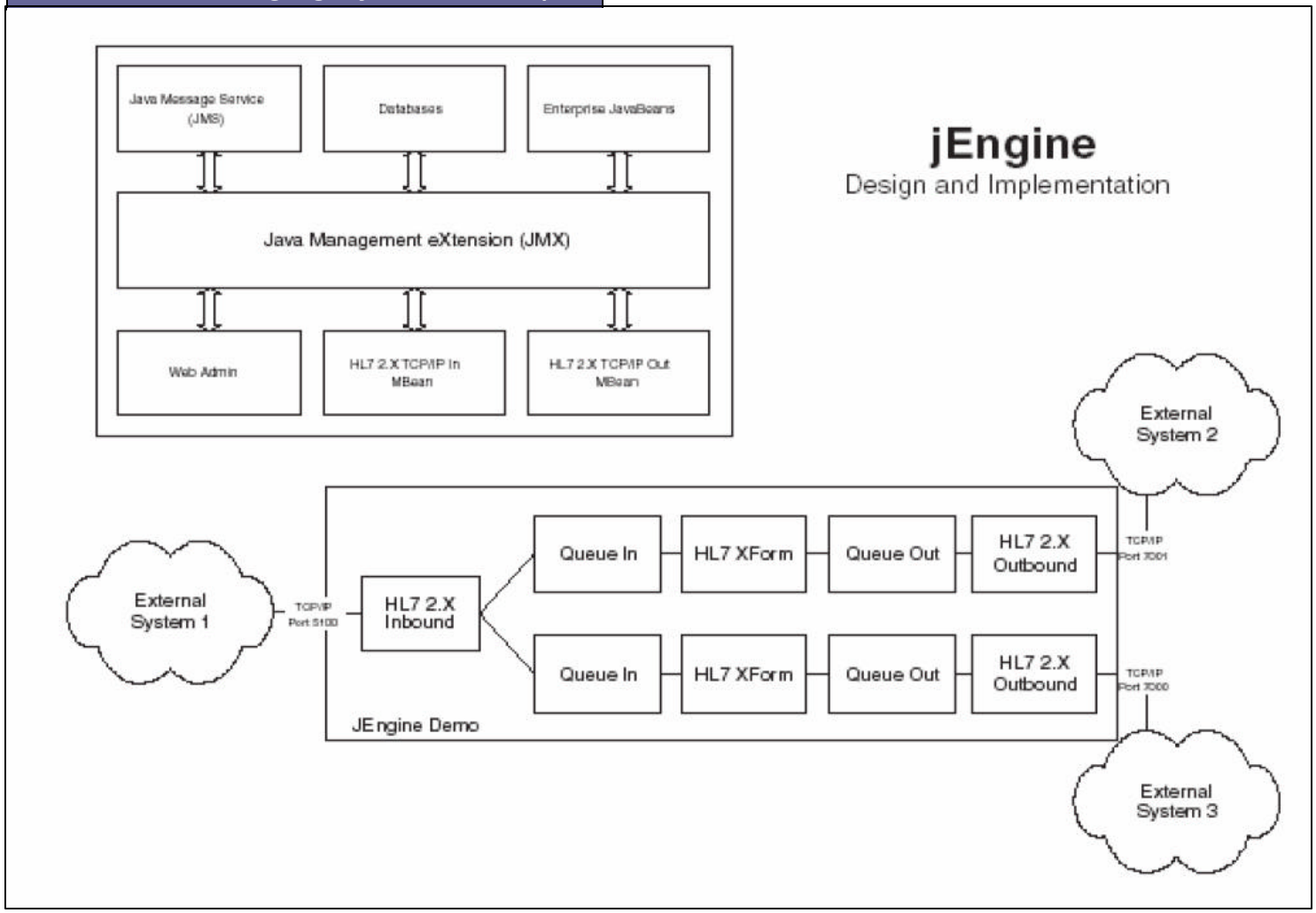


SILVER LEVEL - INTERFACE INTEGRATION

Standard messaging systems where appropriate

Examples: HL-7 Messaging Model

Standard Messaging System Example



SILVER LEVEL – VOCABULARIES/TERMINOLOGIES & ONTOLOGIES

Usage of standard terminologies approved by public standard bodies or caBIG Vocabulary/CDE Workspace

Examples: CIAF Medical Coding Module

Usage of Standard Vocabulary Example

*CIAF Medical Coding Module

```
private MatchedKeywordConceptsVO getConcepts(String keyword, VocabularyVO vocabularyVO) {

    MetaThesaurusConceptSearchCriteria mtcsc = new MetaThesaurusConceptSearchCriteria();
    MetaThesaurusConcept mtc = new MetaThesaurusConcept();
    MatchedKeywordConceptsVO matchedVO = new MatchedKeywordConceptsVO();
    matchedVO.setKeyword(keyword);
    mtcsc.setSearchTerm(keyword);
    logger.debug("Keyword:" + keyword);
    Source src = new Source(); //Set source in the search criteria
    src.setAbbreviation(vocabularyVO.getSourceName());
    mtcsc.setSource(src);
    mtcsc.setShortResult(false);
    mtcsc.setLimit(200);

    //call search
    Concept[] conceptArray = mtc.search(mtcsc);
    FilterChainer chainer = new FilterChainer();

    // Using Chain of Responsibility Patterns to set-up a configurable Filtering mechanism
    // FilterCriteria based on the vocabulary,
    // read from XML configuration file.

    FilterCriteria criteria = new FilterCriteria();
    criteria.setVocabulary(vocabularyVO.getSourceName());
    List matchingConcepts = Arrays.asList(conceptArray);
    chainer.setupChain(criteria);
    chainer.startChain(matchingConcepts, criteria);
```

SILVER LEVEL – DATA ELEMENTS

Use common standard electronic representation for CDE's such as ISO 11179 or comparable standard
Harmonization and re-use of CDEs; CDEs built using standard controlled terminologies approved by caBIG Vocabulary/CDE Workspace
Examples: CIAF accessing caDSR via caBIO API using RMI

Usage of Standard Vocabulary Example

*CIAF caDSR Adapter Code

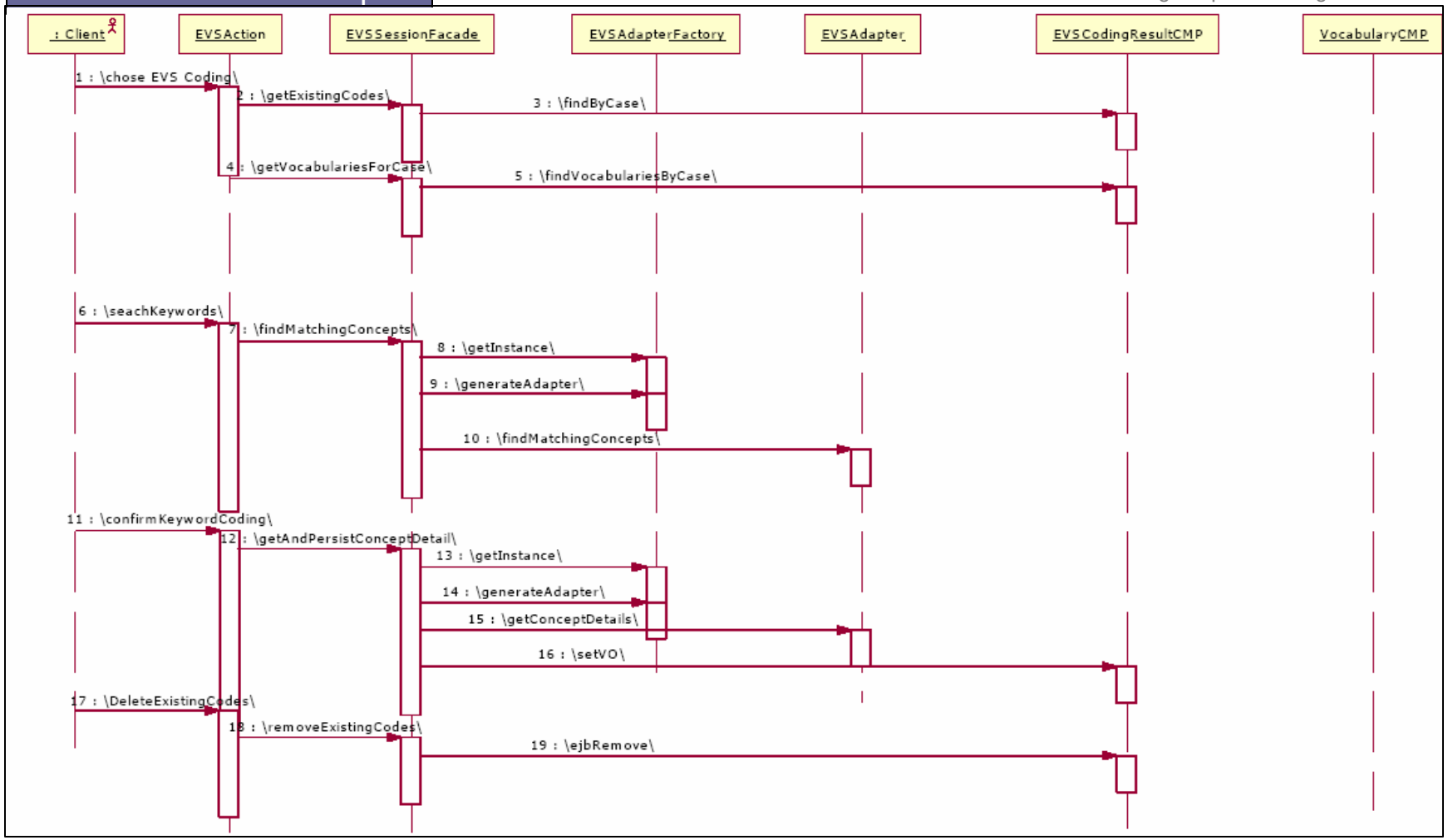
```
public class CDEQuery {
    public CDEVO getCDE(CDEKey CDEId)
    {
        DataElement dataElement = new DataElement();
        DataElementSearchCriteria criteria = new DataElementSearchCriteria();
        SearchResult result = new SearchResult();

        try
        {
            criteria.setPublicId(new Long(CDEId.getPublicId().longValue()));
            result = dataElement.search(criteria);
        }
        // Adapter Layer: Catching all Exceptions, abstracting layers above from any connection failures etc.
        catch(Exception e)
        {
            String message = "Adapter failure getting CDE";
            logger.error(message, e);
            throw new CIAFAdapterException(message, e);
        }

        //Populate the CDEVO and return back to be placed into CDE Cache
        CDEVO cdeVo = new CDEVO();
        // ... populate cdeVo
        return cdevo;
    }
}
```

SILVER LEVEL – INFORMATION MODEL

Use of standard modeling language such as UML to represent the information model
Examples: CSAERS Medical Coding UML diagrams – Sequence Diagram

Information Model Examples

*CSAERS Medical Coding Sequence Diagram

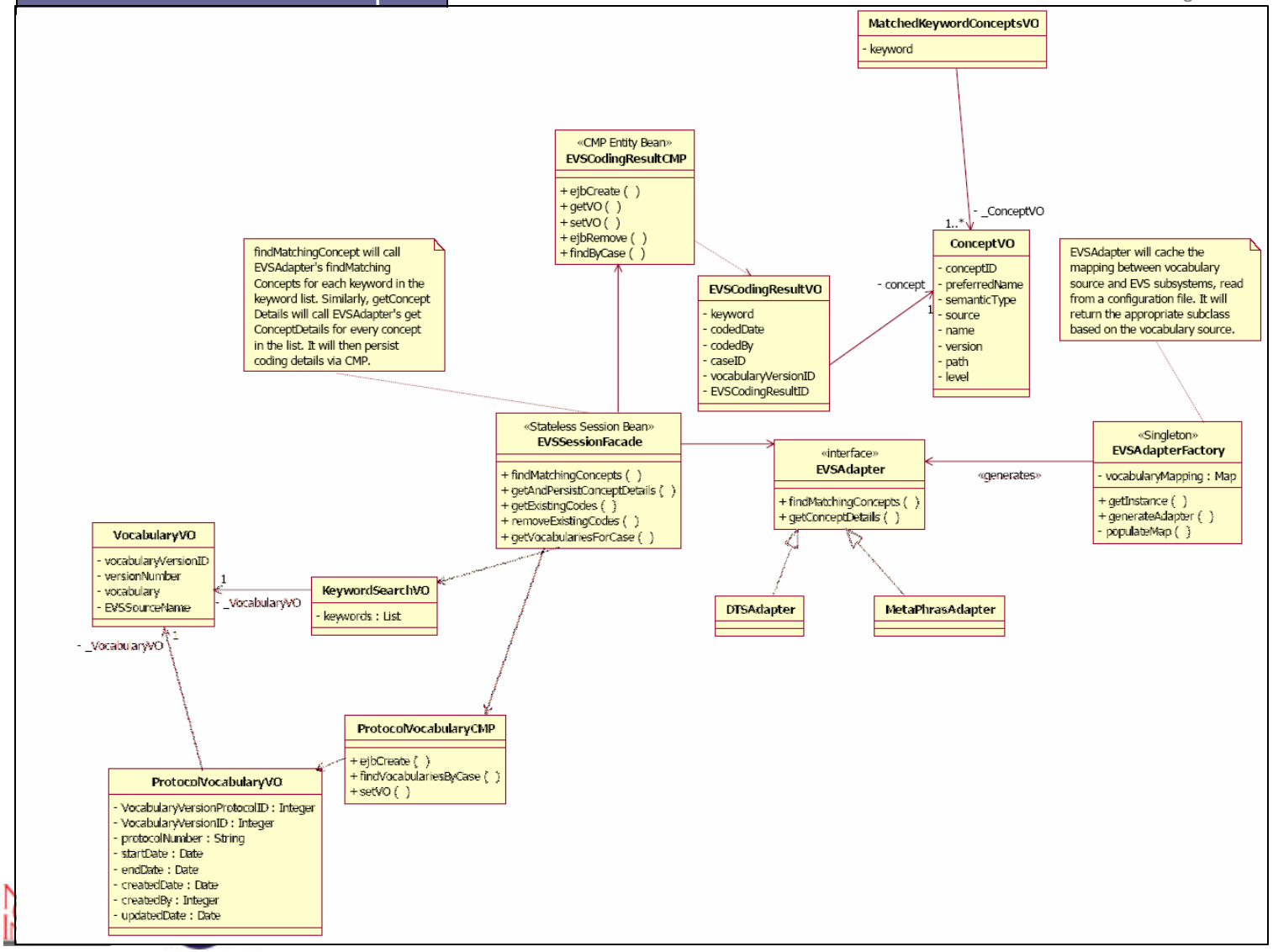


SILVER LEVEL – INFORMATION MODEL

Use of standard modeling language such as UML to represent the information model
Examples: CSAERS Medical Coding Class diagram

Information Model Examples

*CSAERS Medical Coding Class Diagram



Q&A