

ABSTRACT-

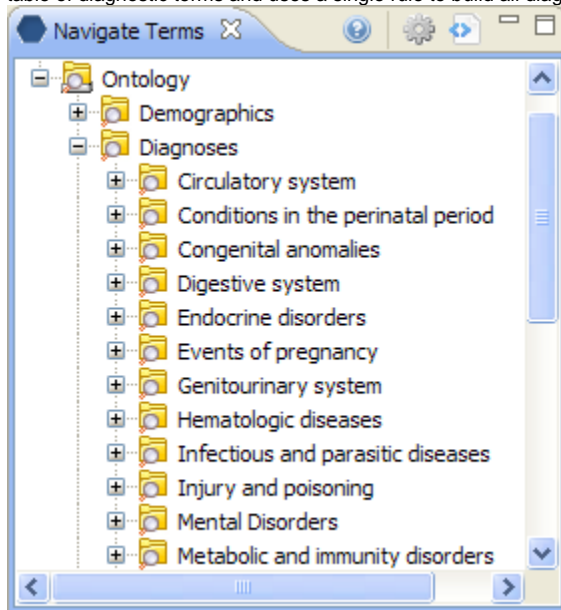
This is a software architecture document for the **Ontology Management (ONT) cell**. It identifies and explains important architectural elements. This document will serve the needs of stake holders to understand system concepts and give a brief summary of the use of the ONT message format.

Overview

The Ontology Management cell (ONT) is an i2b2 Hive core cell. This cell manages i2b2 vocabulary definitions and contains concepts and information about relationships between concepts for the entire hive. It is accessed by other cells to give semantic meaning to data.

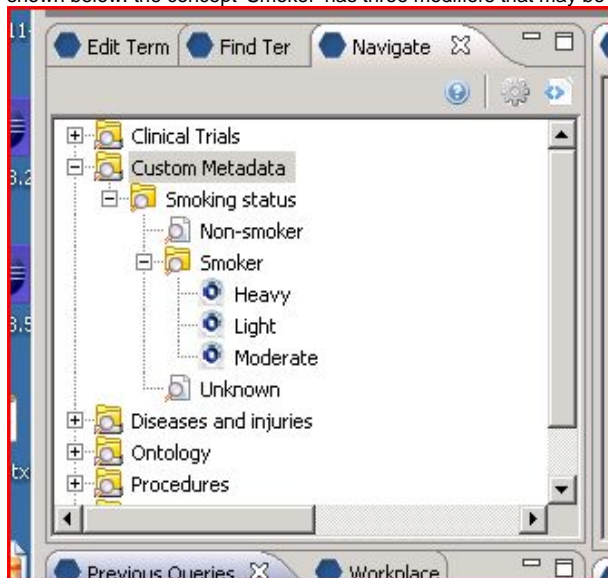
Vocabularies in the ONT cell are organized in hierarchical structures that represent the relationships between terms. The top levels in the hierarchy are called the "**parents**" or "**roots**", with the lower levels being their "**children**". Elements occurring on the same level are known as "**siblings**". A level in a hierarchy is sometimes referred to as a "**node**", and a group of related data is called a "**category**".

A category is defined as a set of data for which there is a common rule or rules for querying against the *Clinical Research Chart (CRC)*. A category is usually represented visually as a table of terms. An example of a category is the "Diagnoses" category shown in the diagram below, which consists of a table of diagnostic terms and uses a single rule to build all diagnostic queries.



Vocabularies in the ONT cell may originate as code from different sources. The ONT cell distinguishes these codes from one another by pre-pending a unique prefix to each code. Each distinct vocabulary and their associated codes are called a **scheme**.

In Release 1.6, we added **modifiers** to the ONT cell vocabulary. A modifier may be used to extend the meaning of a concept. An example of modifiers is shown below: the concept 'Smoker' has three modifiers that may be applied to it: 'Heavy', 'Light', and 'Moderate'.



ONT Definitions, Acronyms and Abbreviations

Vocabulary Data Object (VDO)

This object holds vocabulary definitions and information about the relationships between concepts.

Scheme

Each distinct vocabulary and their associated codes is called a scheme. A distinction is made between codes from different sources by pre-pending a unique prefix to each code.

Roles

When and how data is presented to a user is based on their user roles, which are specified in the PM Cell. Each user will have at least two roles per user_id and product_id combination. These two roles can be further defined as a Data Protection role and a Hive Management role. The data protection role establishes the detail of data the user can see while the hive management role defines their level of functionality the user has in a project. The following tables summarize the roles in a hierarchical order of least to most access.

Data Protection Track	
Role	Access Description
DATA_OBFSC	<p>OBFSC = Obfuscated</p> <ul style="list-style-type: none"> The user can see aggregated results that are obfuscated (example: patient count). The user is limited on the number of times they can run the same query within a specified time period. If the user exceeds the maximum number of times then their account will be locked and only the Admin user can unlock it.
DATA_AGG	<p>AGG = Aggregated</p> <ul style="list-style-type: none"> The user can see aggregated results like the patient count. The results are <u>not</u> obfuscated and the user is <u>not</u> limited to the number of times they can run the same query.
DATA_LDS	<p>LDS = Limited Data Set</p> <ul style="list-style-type: none"> The user can see all fields except for those that are encrypted. An example of an encrypted field is the <i>blob fields</i> in the <i>fact</i> and <i>dimension tables</i>.
DATA_DEID	<p>DEID = De-identified Data</p> <ul style="list-style-type: none"> The user can see all fields including those that are encrypted. An example of an encrypted field is the <i>blob fields</i> in the <i>fact</i> and <i>dimension tables</i>.
DATA_PROT	<p>PROT = Protected</p> <ul style="list-style-type: none"> The user can see all data, including the identified data that resides in the Identity Management Cell.

Hive Management Track	
Role	Access Description
USER	Can create queries and access them if he/she is the owner of the query.
EDITOR	Can create, delete or edit Ontology metadata terms
MANAGER	Can create queries and can access queries created by different users within the project.
ADMIN	

- Further details regarding roles can be found in the PM_Design_Document.*

Security

Users may access ONT with a user-id and password combination, which is authorized through the Project Management Cell. The implementation detail of Project Management Cell is considered out-of scope to this document.

Scope of the system

Some other participants, currently outside the scope of ONT are:

- Project Management Cell

Assumptions/Constraints

The Ontology metadata database shall not contain protected health information.

Technical Platform

The product is designed to run on the following platform:

- Java 2 Standard Edition 7.0
- Oracle Server 10g/11g database
- SQL Server 2005/2008
- Xerces2 XML parser
- Jboss Application server version 7.1.1
- Spring Web Framework 2.0
- Axis2 1.6.2 web service (SOAP / REST)

Transaction

The ONT system is transactional, leveraging the transaction management model of the J2EE platform.

Security

The application must implement basic security behaviors:

- Authentication: Authenticate using at least a user name and a password
- Authorization: User may only access categories that they are allowed to by role
- Confidentiality: Sensitive data must be encrypted
- Data integrity : Data sent across the network cannot be modified by a tier
- Auditing: In the later releases we may implement logging of sensitive actions

Persistence

This application utilizes JDBC calls to retrieve persisted data.

Reliability/Availability

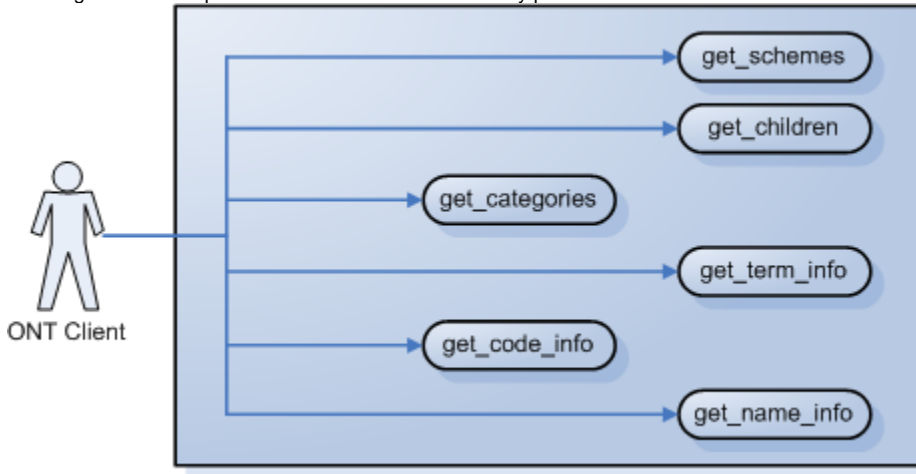
- The Reliability/Availability will be addressed through the J2EE platform
- Targeted availability is 16/7: 16 hours a day, 7 days a week
- The time left (8 hours) is reserved for any maintenance activities

Performance

The user authentication with project management cell must be under 1 second.

Use Case

The diagram below depicts common use cases a user may perform with the ONT cell.



Operations

The ONT service is designed as a collection of operations or use cases:

Service Operation	Description
getCategories	Returns a list of categories available for a given user. These categories are displayed in a tree format. The top level of the tree consists of all the categories a particular user has permission to see.
getChildren	Expands any level of a vocabulary category, providing information about its children, for a given user.
getSchemes	Returns a list of schemes available in the system. This operation basically provides information about the different kinds of coding systems that exist.

getNameInfo	Returns information needed about all nodes related to a given search keyword or name.
getCodeInfo	Returns information about a code, such as the name associated with a particular code.
getTermInfo	Returns information about a particular node.
addChild	Adds a new Ontology term to the tree under the selected parent node.
deleteChild	Deletes a selected Ontology term.
modifyChild	Modifies content within an existing Ontology term
updateCRCConcept	Notifies the Ontology cell to synchronize metadata terms with concept_dimension table
getProcessStatus	Returns status information about the concept_dimension synchronization process
getDirtyState	Returns state information about the need to synchronize with the concept_dimension table
updateConceptTotalNum	Notifies the Ontology cell to get patient count from CRC for this concept and then update the totalnum for this concept in the metadata table.
getModifiers	Returns list of modifiers for a concept if they exist.
getModifierChildren	Expands any level of a modifier folder, providing information about its children.
getModifierInfo	Returns information about a particular modifier.
getModifierNameInfo	Returns modifiers associated with a concept that meet a name search criteria
getModifierCodeInfo	Returns modifiers associated with a concept that meet a code search criteria.
addModifier	Adds a new Ontology modifier to the tree under the selected parent node or modifier.
excludeModifier	Excludes an existing modifier from a concept lower in the hierarchy than the modifier's specified hierarchy level.

Architecture Description

This section provides a description of the architecture as multiple views. Each view conveys the different attributes of the architecture.

1. Components and Connector View
2. Client-Server Style

1. Module View
2. Decomposition Style
3. Uses Style

1. Data View
2. Deployment View

Components and Connector View

A **Component and Connector view** represents the runtime instances and the protocols of connection between the instances. The connectors represent the properties such as concurrency, protocols and information flows. The diagram shown in the *Primary Presentation* section represents the Component and Connector view for the multi-user installation. As seen in the diagram, component instances are shown in more detail with specific connectors drawn in different notations.

Client-Server Style

The ONT system is represented using the components and connector client-server view.

Primary Presentation

Metadata*i2b2*
 WorkbenchOntology
 Management
 ServerProject
 Management
 ServerREST/SOAPREST/SOAPJDBC*Server*Client

3.1.1.2 Element Catalog

Element Name	Type	Description
i2b2 Workbench	Client Component	Webservice client submits the requests to ONT Server components and renders response XML.
Ontology Management Server	Server Component	Provides Web Service Interface for the ONT system. It supports the REST or SOAP protocol. It directs the user to the correct data source associated with the project. It uses Project Management server to handle user authentication.
Project Management Server	Server Component	ONT cell uses Project Management cell to authenticate user. ONT cell constructs PM request message and makes a web service call to Project Management Cell.
Metadata	Data Repository Component	This repository is a database for i2b2 metadata.
JDBC	Query Connector	SQL query used as a connector between the ONT System and the Metadata database.
Web Service	Request Connector	REST protocol used to communicate with the external system.

Design Rationale, Constraints

N-tier Architecture

The client-server style depicts an n-tier architecture that separates the presentation layer from the business logic and data access layer.

Module View Type

The module view shows how the system is decomposed into implementation units and how the functionality is allocated to these units. The layers show how modules are encapsulated and structured. The layers represent the "*allowed-to-use*" relation.

The following sections describe the module view using Decomposition and Uses Styles.

Decomposition Style

The Decomposition style presents system functionality in terms of manageable work pieces. It identifies modules and breaks them down into sub-modules and so on, until a desired level of granularity is achieved.
 Primary Presentation

System	Segment
Ontology Management Server	Operation Manager

Element Catalog

Element Name	Type	Description
Operation Manager	Subsystem	This subsystem manages queries for ontology operations.

Context Diagram

Operation ManagerOntology Management Server

Uses Style

The Uses style shows the relationships between modules and sub-modules. This view is very helpful for implementing, integrating and testing the system.
 Primary Presentation

System	Segment
Ontology Management Server	ONT Module
Operation Manager Subsystem	Ontology Webservice
	Request Handler
	Request DAO
	Vocabulary Data Object

Element Catalog

Element Name	Type	Description
ONT Module	Module	Authenticates user through PM Server System
Ontology Webservice	Communication Module	Provides web service interface to ontology operations.
Request Handler	Business Object	Delegates Ontology requests to Data Access Object layer to perform database operations.
Request DAO	Data Access Object	Supports database query operations.
Vocabulary Data Object	Transfer Object	Object representation of persisted data