Virginia Tech

VLDS Lexicon Metadata Tool

Design Document

Aaron Schroeder
10/23/2012

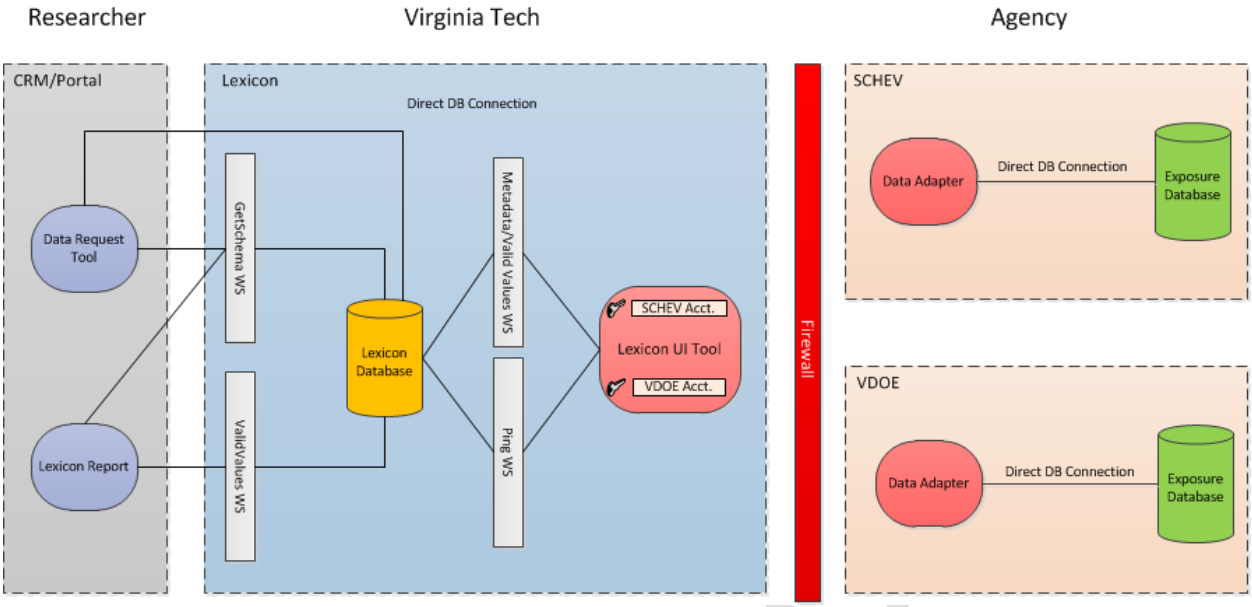
Contents

Version #	Date	Entered by	Comments
1.0	6/1/2012	Aaron Schroeder	Initial Documentation and overview of metadata import process
1.1	6/8/2012	Aaron Schroeder	Added Functional Requirements
1.2	6/11/2012	Aaron Schroeder	Added Lexicon Overview Diagram, log data section, web services section, small updates sub section, and embedded database section
1.3	6/14/2012	Austin Mills	Updated Lexicon Overview Diagram
2.0	6/20/2012	Aaron Schroeder	Added Lexicon Database structure
2.1	6/22/2012	Aaron Schroeder	Added LMT user roles section
2.2	7/17/2012	Austin Mills	Added Installation section

Overview

The Lexicon Metadata Tool (LMT) is an interface built in the Lexicon environment that has the ability to easily modify/create/import metadata and view Lexicon system connections, logs, and statistics for each agency in the VLDS. Although the tool is not hosted at the agency level, each agency will have access to its own installation.

LEXICON OVERVIEW



Functional Requirements

Req-1	Account privileges divided by admin, read/write, read-only, and audit
Req-2	Contain web service to test Lexicon Connection
Req-3	Must be able to gather Lexicon statistics through a web service
Req-4	Log system usage per user for auditing purposes
Req-5	Must have the ability to import batch file for mass updates
Req-6	Must have the ability to make metadata and valid value changes
Req-7	Interface with Lexicon databases residing on the following types of databases: ?
Req-8	Must install on the following operating systems: ?

LMT Installation

This section describes the requirements and procedures for installing the Lexicon Metadata Tool for an agency.

Requirements

In order to install the Data Adapter the newest version of Netbeans should be installed along with a command line subversion program.

Netbeans 7.1.x: http://netbeans.org/downloads/index.html

Command line SVN (recommended install): http://subversion.apache.org/

Procedure

- Check out the repository: http://173.203.76.91:8989/svn/vt_vlds/LexiconMetadataTool
- Start Netbeans and open the project from the *trunk* folder in the newly checked out tree: *<svn-root>/trunk/LexiGui*
- In Netbean's *Projects* tab, right click on the *LexiGui* project you just opened. Select: *Custom | Build Install Package* from the dropdown menu.
- The build package is placed in a folder in *<svn-root>/trunk/LexiGui/* It has the name *LexiconUI_XXX?* Where XXX is the version number built. If any modifications have been made to the code before building the install package there will be an 'M' after the build number. This signifies an un-official custom build.
- To install/run the Lexicon Metadata Tool, run the *startup.bat* file at the top level of the folder.
- This should be zipped up before release – this can be done with whichever zip utility you like.

Check the Data Adapter's version

To verify which version of Data Adapter you're running:

- Logon to the Data Adapter web interface: https://localhost:8080/VLDS/
- Select the *Home* tab on the left. Find the ***LexiconUI Version ID*** line. Note: If it is a developer's custom build, the build number will be followed by this warning: *(This is an un-official developer's version)*

Web Services

The Lexicon UI Tool interacts directly with web services to communicate with the Lexicon Database for metadata/valid values management and testing the connection. There are no other integration points for the UI Tool outside of these web services.

Ping Web Service

This simple web service is used to test the connectivity between the Lexicon Database and the Lexicon UI Tool. The response should have HTTP status code of 200 (OK) with a text/plain content type.

Metadata Web Service

This web service is used when the Lexicon Database queries the UI Tool for extended properties metadata.

Valid Values Web Service

This web service is used when the Lexicon Database queries the UI Tool for a specific Valid Values column or all Valid Values.

Metadata/Valid Values Management

The Lexicon Metadata Tool's main function will be to manage VLDS metadata and valid values for each agency. Modifications can be done through the user interface or through the import tool built into the software which allows for batch uploads. Read/Write users can use whichever option better suits their needs for metadata and valid values management. The two update processes are explained below.

Small Updates

Metadata and Valid Values can quickly be created, modified, or deleted using the Lexicon user interface. This interface is best suited for small updates that can be done quickly by the agency. To perform mass updates, the import tool should be used.

Mass Updates

The Lexicon import tool will allow the user to upload mass metadata and valid value changes by uploading a file. The file type(s) is/are currently undecided and will most likely be a backup file or SQL script file, which would be generated by various tools used at each agency. The code design in the LMT allows for easy extension to different databases by breaking the job into two parts. The first part is to re-create the tables in a temp database from the given file and secondly the code will generate the change list for approval and audit. To extend the tool, it will only be necessary to implement step one.

Process

The procedure to mass update the Metadata and Valid Value tables are as follows: The user will first make appropriate changes locally to their local Metadata and Valid Values table(s) followed by exporting these tables to an acceptable file type. Users will then login with their credentials to the LMT and select the Import tab. After clicking browse and uploading the local exported file, a list of differences will be generated and displayed. The user has several filters available to check the data and ensure the changes are correct. After approving the changes, they will be submitted and added to the audit log.

Log Data

The Lexicon Metadata Tool logs each action and event to a log table. These events can easily be searched for and displayed using the user interface. The table below gives the table structure for the log data.

Column Name	Data Type	Description
LOG_DATE	DATETIME	Date log was generated.
LOG_LEVEL	VARCHAR(20)	Log level (e.g. 'TRACE','DEBUG','INFO','WARN','ERROR','FATAL').
THREAD	VARCHAR(50)	Name of the Java thread that generated the logging event.
CATEGORY	VARCHAR(100)	Category of the log event (typically a Java class name).
MESSAGE	VARCHAR(2000)	The log message.

Embedded Database

The Lexicon Metadata Tool will have an embedded database associated with it similar to the Data Adapter. This database will only hold data regarding agencies, configurations, web service stats, and other system information. The database will not be used to hold metadata or valid value tables.

Lexicon Database Metadata Tables

Each participating agency will have their metadata and valid value tables stored in the Lexicon Database. The LMT (Lexicon Metadata Tool) is built to manage these tables in the Lexicon. The VALID VALUES, TABLE_METADATA, and COLUMN_METADATA tables are used as the permanent location for this data.

VALID_VALUES Table Structure

This table is used to store valid values for columns in exposure database data sets. This data can be managed using the LMT manually or with the import tool. This data permanently resides in the Lexicon which in turn provides it to the DRT (Data Request Tool) and Lexicon Report. When researchers are composing a query using the DRT they provide a WHERE clause with filters. The researcher can then choose a value from the valid values of a filter element.

Column name	Data type	Description
TABLE_NAME	VARCHAR(100)	This is the name of a table in the Exposure-DB.
COLUMN_NAME	VARCHAR(100)	
VALUE	VARCHAR(500)	Dates must be formatted as 'yyyy-mm-dd'.
DESCRIPTION	VARCHAR(2000)	This will be displayed to the researcher along with the value. It will not be displayed in the query result set.
VALID_USE_BEGIN_DATE	DATE	Required
VALID_USE_END_DATE	DATE	
LAST_UPDATE	TIMESTAMP (as defined in SQL-92 standard, or closest data type)	For auditing purposes only.

Valid Value Notes:

- Integer and decimal ranges can be specified for column data. A column with an integer or decimal range will have one record with the start of the range and one record with the end of the range. For example, we specify a range of 0..45, 50, 55..100 by splitting the begin and end numbers of the range and designate them to a single row. The first record will contain the value

'0' and the next record will contain the value '45'. For more examples please see the example table below.

- The Description field is a description of a *Value*. Note that COLUMN_METADATA.MS_DESCRIPTION is the proper data element for identifying a *Column*'s description to the Lexicon. When a valid value or a range of valid values are entered, a thorough description must be entered.
- Dates entered as valid values can also be specified as a range. The format used for dates should be 'yyyy-mm-dd'.For example '2012-05-15' is a proper date format. Fields that can contain any kind of free form text should not have valid values listed in the table. It should only be explained in the column metadata table with a detailed description. The same can be applied to strings with a specific format.
- If Valid Values have the ability to contain a "Null" value, it should be specified by using the text string "VLDS_NULL." The Lexicon will expose VLDS_NULL values as database nulls in both the Lexicon Views and web service. This should not be placed in any other column or table except the Valid Values table as a possible value. Exposed data sets should not have this value entered the tables.

Example VALID_VALUES data:

TABLE_NAME	COLUMN_NAME	VALUE	DESCRIPTION
Building	RoomNum	100	VLDS_RANGE_BEGIN. Beginning of the range of rooms 100-120 in the first floor of the building.
Building	RoomNum	120	VLDS_RANGE_END. Ending room number for the range 100-120.
Building	RoomNum	125	First floor auditorium.
Building	RoomNum	130	VLDS_RANGE_BEGIN. Beginning of the range of rooms 130-150 in the building.
Building	RoomNum	150	VLDS_RANGE_END. Ending room number for the range 130-150.
Building	RoomNum	VLDS_NULL	Null value
Person	Gender	M	Male
Person	Gender	F	Female
Course	Semesters_Offered	01/15/2012	VLDS_RANGE_BEGIN. Start of the spring semester date range.
Course	Semesters_Offered	05/15/2012	VLDS_RANGE_END. End of the spring semester

			date range. Course not offered in summer.
Course	Semesters_Offered	08/15/2012	VLDS_RANGE_BEGIN. Start of the fall semester date range. Course not offered in summer.
Course	Semesters_Offered	12/10/2012	VLDS_RANGE_END. End of the fall semester date range.
Chemistry	ph_Scale	0.0	VLDS_RANGE_BEGIN. Beginning scale of acidity or basicity rounded to only one decimal place.
Chemistry	ph_Scale	14.0	VLDS_RANGE_END. Ending scale of acidity or basicity rounded to only one decimal place.

TABLE_METADATA Table Structure

The TABLE_METADATA table is used to store metadata for agency exposure database tables. This data can be managed using the LMT manually or with the import tool. This data permanently resides in the Lexicon which in turn provides it to the DRT (Data Request Tool) and Lexicon Report.

Note that Views in the Lexicon can only have a 1-to-1 relationship to Exposure-DB tables.

Column name	Data type	Description
TABLE_NAME	VARCHAR(100)	This is the name of a table in the Exposure-DB.
FRIENDLY_NAME	VARCHAR(500)	
MS_DESCRIPTION	VARCHAR(2000)	
CRITICAL_CHANGES	VARCHAR(4000)	
LAST_UPDATE	TIMESTAMP (as defined in SQL-92 standard, or closest data type)	For auditing purposes only.

COLUMN_METADATA Table Structure

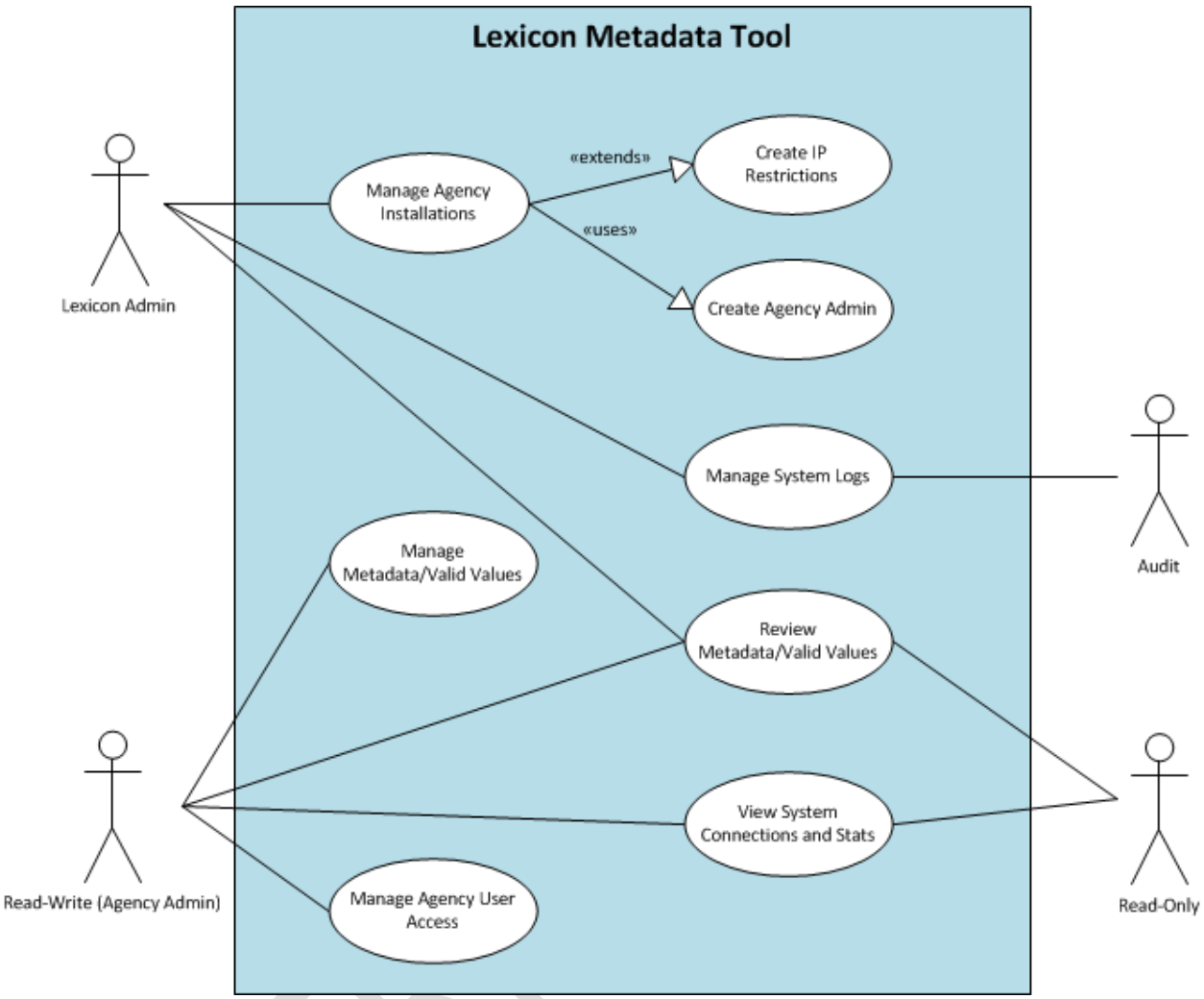
The COLUMN_METADATA table is used to store metadata for agency exposure database columns in a table. This data can be managed using the LMT manually or with the import tool. This data permanently resides in the Lexicon which in turn provides it to the DRT (Data Request Tool) and Lexicon Report.

Note that Views in the Lexicon can only have a 1-to-1 relationship to Exposure-DB tables.

Column name	Data type	Description
TABLE_NAME	VARCHAR(100)	This is the name of a table in the Exposure-DB.
COLUMN_NAME	VARCHAR(100)	
FRIENDLY_NAME	VARCHAR(500)	
MS_DESCRIPTION	VARCHAR(2000)	
CRITICAL_CHANGES	VARCHAR(4000)	
DATA_DOMAIN	VARCHAR(200)	Comma separated
JOIN_ONLY	CHAR(5)	true/false
VALID_USE_BEGIN_DATE	DATE	Required
VALID_USE_END_DATE	DATE	
LAST_UPDATE	TIMESTAMP (as defined in SQL-92 standard, or closest data type)	For auditing purposes only.

LMT User Roles

In the Lexicon Metadata Tool, there are four types of users: Lexicon Admin, Audit, Read-Write, and Read-Only. Each role has certain privileges and access to the LMT. The diagram below shows each user and the actions they can take.



LMT Admin

The Admin role is the LMT administrator who has no agency affiliation and is in charge of installing the LMT for each agency. Similar to the Audit user, the Admin can manage the system logs for each agency. The Admin can also restrict users to the LMT by IP ranges to help prevent security threats. They don't have the ability to manage metadata for each agency using the LMT, but have Read-Only access. One important responsibility of the Admin is to create a Read-Write (Agency Admin) user for the agency after installing the LMT for that agency.

Audit

This user is on an agency level and has the ability to manage LMT log files to keep track of all system changes including metadata/valid values updates. The Audit user does not have any Read-Write or

Admin privileges. This role is designation for a specialized user who has similar responsibilities as the Read-Write user.

Read-Write (Agency Admin)

The Read-Write user is also referred as the Agency Admin and has the highest level of access for the agency installation done by the LMT Admin. This user has the ability to make manual metadata changes as well as use the import tool for batch changes. They also have the ability to create Read-Only and Audit users for their agency.

Read-Only

The Read-Only user has limited abilities in the LMT. The primary purpose of this user is to view all the metadata and valid values without making any changes. They also have the ability to see system connections and statistics. This account could also be used for demo purposes or any situation where the user doesn't want to accidently change any records.