

2015.12 Technical Face-To-Face Meeting Notes

Document Information
Author: Craig Stancl, Scott Bauer, Cory Endle Email: Stancl.craig@mayo.edu, bauer.scott@mayo.edu, endle.cory@mayo.edu Team: LexEVS Contract: S13-500 MOD4 Client: NCI CBIIT National Institutes of Health US Department of Health and Human Services

Contents of this Page
• Tuesday, December 8, 2015 ◦ 9:00 AM - 9:30 AM ◦ 9:30 AM - Noon ▪ Discussion: Tech Stack Updates ▪ Discussion: LEXEVS API/Browser Performance and Usability Improvements ◦ 1:00 PM - 2:00 PM ◦ 2:00 PM - 4:00 PM ▪ Discussion: Enhancing the LEXEVS CTS2 REST Interface ▪ Information Session: Using LexEVS Query API and CTS2 API ▪ Next Steps • Wednesday, December 9, 2015 ◦ 8:00 AM - 9:00 AM ◦ 9:00 AM - 9:30 AM ◦ 9:30AM - Noon ▪ Discussion: Coding Scheme Search and Indexing ▪ Discussion: LEXEVS Loader Improvements ◦ 1:00 PM - 1:30 PM ◦ 1:30 PM - 4:00 PM • Thursday, December 10, 2015 ◦ 9:00 AM - 9:30 AM ◦ 9:30 AM - Noon ▪ Discussion: Triple Store/Graph Database ▪ Discussion: Cloud Considerations and Discussion: Build and Deployment Process ◦ 1:00 PM - 3:00 PM ▪ Discussion: Build and Deployment Process and Discussion: Value Set Editor (Authoring) ▪ Discussion: The Future of IbGUI ◦ 3:00 PM - 5:00 PM ▪ Debrief

The **purpose of this document** is to document the technical face to face meeting details between the NCI and Mayo for the National Cancer Institute Center for Biomedical Informatics and Information Technology (NCI CBIIT) **LexEVS Release 6.3** and **LexEVS Release 6.4** .

2015 December Face-to-Face Meeting Notes

Tuesday, December 8, 2015

9:00 AM - 9:30 AM	1W030	Overview and Planning
-------------------	-------	-----------------------

Attendees: Jacob, Larry, Kim, Jason, Craig, Cory, Scott, Sarah Elkins, Gilberto, Rob

- The group spent time reviewing the proposed agenda to understand who will need to attend and other logistics for the day (Tuesday)
- 9:30 - Noon
 - Tech stack was discussed to determine which aspects of the tech stack to cover. First, cover the items we have listed. .
 - LexEVS API Browser section - both EVS and LexEVS development teams.
 - It was noticed that it is now a tech catalog - and not tech stack.
- 1:00 - 4:00 will include CTRP group
 - It was noted that they will move away from Oracle.
 - Need to discuss MySQL
- 4:00 - 5:00
- Thursday - Sarah may try to attend the graph/DB session.

- Automating Data Loads and customization to be included in the Wednesday 1:30-4:00.

9:30 AM - Noon	1W030	Discussion: Tech Stack Updates - Migration to Java 8 - Migration to CentOS 7 - Migration to Spring 4 - Determine next steps/roadmap
		Discussion: LEXEVS API/Browser Performance and Usability Improvements - NCI to provide requirements/use cases - Areas for discussion: - value set - mapping - multi-namespace - relationship querying - inferred data - Discuss the removal of caCORE and refactor Remote API - Determine next steps/roadmap

Attendees: Jacob, Larry, Kim, Jason, Craig, Cory, Scott, Sarah Elkins, Gilberto, Rob, Sherri

Discussion: Tech Stack Updates

- centOS 7
 - centOS 7 hasn't been used by the Mayo team (as of yet.)
 - There is much interest in the automated environment setup provided by Docker.
 - Sarah indicated that it would **NOT be an upgrade on the existing web servers**. They would recommend this for new servers.
 - As of Sept 30, CentOS 7 is to the current supported catalog at NCI**. They still provide previous CentOS 6 for several years. Redhat dates 2020 for CentOS 6 and 2024 for CentOS7.
 - Scott also suggested the use of a Sonatype Nexus server to distribute Docker artifacts. NCI currently provides a Nexus server. However Docker artifacts are fairly large and infrastructure would need to be considered.
 - Additional Docker discussion later in the week. There are 2 use cases - deploying at NCI and supporting other users (setup).
 - MariaDB is a MySQL fork. Advantages aren't known at this time. The NCI database team would need provide support if MariaDB were to be used. This may be in the future Tech Catalog.
 - Sarah recommended that we hold off on deciding on CentOS 7 until the use of Docker is further defined.**
 - Larry pointed out that CentOS 7 has additional security enhancements. Jacob indicated there is no current mandates to move to CentOS 7.
 - Gilberto shared that the Protege Project is planning on using Docker.
- Java 8
 - Scott shared that he was able to compile LexEVS with Java 8, but no testing had been completed.
 - Scott looked at the potential Tomcat issue and determined that this is not a concern if using Tomcat 7.0.5.8 or higher.
 - Gilberto shared that for the Protege Project uses Java 8 - the only issues were around concurrency.
 - Sarah indicated that there are other projects that are built with Java 8 and Java 8 and Tomcat 7. There is full support for Java 8 and Tomcat 7 at NCI. Sarah would need to know the exact version numbers.
 - Kim can demonstrate an error with Java 8 and Spring 2.7. Kim and Scott to discuss.
 - Gilberto indicated that will need to move to Java 8. Oracle dates are what drive moving from version to version.
 - There is no plan to deprecate Java 7 at NCI** (typically a 2 year lead time).
 - Tomcat 7 Java 1.7 will be deprecated in 2017.
 - This will be considered for 6.5.**
- Spring Migration
 - Scott indicated that after testing the Spring 4 - there are much DAO changes.
 - This should be considered with Java 8 upgrade.**
- Roadmap/Next Steps
 - centOS 7 - no immediate plans** - defined after further discussions around use of Docker.
 - Will need to consider hardware configurations to support CentOS 7 deployments (March 2016 timeframe)
 - Updates considered for 6.5 timeframe:**
 - Java 8
 - Spring 4

Discussion: LEXEVS API/Browser Performance and Usability Improvements

- Value Set
 - Rob noted that using the compiled value set definitions requires an extra step. Tracy indicated that they aren't sure how those are being used.
 - Need to look at compiled value set definitions.**
- Tree Performance
 - Kim noted that it would be good to provide capability to identify if a node is a leaf node. Can currently do this with multiple calls.
 - Scott noted there are a couple of pending JIRA items; notably, the iterator issue - knowing the graph size (provide number of nodes)
 - Kim noted that Triple Store usage - GLEEN - provides extension for graph query in triple store.
 - GLEEN <http://sig.biostr.washington.edu/projects/ontviews/gleen/>
- Usability

- Kim suggested way to provide inbound and outbound
 - Scott noted that we shouldn't change existing API, but add extensions.
 - Kim indicated that current model is complex for the API - requires multiple calls to query.
 - Scott requested that a list of API enhancement requests be provided by Kim and then we can prioritize.
 - Larry indicated that there has been a much larger usage of the CTS2 REST API, so it may make sense to focus on CTS2 and not Java API.
 - Kim - CTS2 API needs to be expanded to fully support browser development. Scott identified that CTS2 implementation would need to be further implemented to support all CTS2 functionality.
 - Kim - LexEVS currently doesn't provide way to prevent pulling entire code system. Scott suggested that the API could govern this functionality.
 - Kim - No way to retrieve concept with role group (class expression in OWL with specific format). Scott noted that this may be impossible.
 - Gilberto - Use case is for the browser - wanting to show role groups. Currently everything is jumbled together.
 - Scott indicated that to support this, a model change would be needed in order to reconstruct this content.
 - Gilberto noted that LexEVS model was to flatten the models and LexEVS wasn't intended to represent this complex modeling. Noted that we need to consider performance. We could also use something to supplement LexEVS terminology services (triple store and LexEVS hybrid)
- Modeling
 - Scott - how can we derive a solution? Is this sufficient within our group or do we reach out to the larger community. Larry indicated that it may be more of a decision for NCI and Mayo group. Scott suggested that if we start down the path of updating the model, we would then like to have additional resources participate.
 - Larry - **identify the use cases around OWL** and then identify resources to review.
 - **It was suggested that we involve Harold to start this discussion.**
 - Gilberto suggested that if we are going to support Round-tripping of OWL and other OWL is to maintain validity of format - this would allow NO loss in and out of LexEVS.
 - Scott indicated this would be a huge effort.
 - Rob noted that LexGrid XML format was also available.
- Multi-Namespaces
 - Gilberto described "Hierarchy traversal based off of namespace" and the other to be able to "Traverse horizontally"
 - Described the use case - Load NCIT and have GO - and crossing namespaces to support crossing coding schemes to GO.
 - Maps would be needed to support crossing coding schemes.
 - Scott noted that for example, with RRF, we already of the CUIs for traversal.
 - Tracy suggested query a coding scheme for it's namespaces and query the other coding schemes for their namespaces.
 - Gilberto noted that OBI may refer to 20 other namespaces - and then have LexEVS determine if those other namespaces exist.
 - **Sherri suggested that the specific use case be documented by Gilberto.**
 - Tracy suggested the use of the URI resolver. Scott indicated that would be overkill for this purpose, but could be used - but noted that it might be heavy weight to have another service.
 - Sherri asked if the identifiers were standard. Gilberto described how identifiers are created - ie. hash.
- Relationship query
 - Gilberto described the use case - For any vocabulary we have - "give me all the concepts that have this association y"
 - Scott noted that this would not be easy in LexEVS. Perhaps SPARQL would be the better choice.
 - Gilberto suggested to flatten out class expression
 - Tracy - anonymous classes are structured - is it possible to create an extension to parse? Scott indicated this is possible - however, performance may be a consideration. It was noticed this would typically be done one at a time.
 - Scott - if flattened, the parenthetical is lost. Gilberto noted we still need to be able to do this.
 - Larry suggested the use of a flag to indicated flattened.
 - Gilberto noted the browser would like to be able to display anonymous classes
 - **Scott suggested we need to determine a specific set of requirements an use cases.**
- Inferred Data
 - Gilberto described the use case - in some areas of vocabulary there aren't specified relationships, but instead they are inferred.
 - For example, Tell me all the organs in the chest cavity - includes heart, even though it's not specified.
 - Tracy described the need to be able to search on them, but my not need/want to see them.
 - Gilberto noted that currently in LexEVS, cannot flag those that are inferred.
 - **This will require additional discussion.**
- Removal of caCORE
 - Scott noted that we've brought this up previously.
 - Web services as part of LexEVS in caCORE framework. These have been around for quite some time.
 - Gilberto indicated that caCORE is no longer being used, so indeed this can be removed.
 - It was noted that users were considered, but it was acknowledged that perhaps we didn't document.
 - **It was decided that caCORE API can be removed.**
- Refactoring Remote API
 - Scott indicated this would all us to update to a more modern Remote API and removal of artifacts from caCORE
 - **This should be considered as part of the Spring migration.**
- Next Steps
 - Use cases can span many of these areas and will to be considered during ongoing discussions.

1:00 PM - 2:00 PM	4W030	Discussion: Enhancing the LEXEVS CTS2 REST Interface • NCI to provide requirements/use cases • Areas for discussion: ◦ Expanding to include Associations ◦ Specialized REST calls (e.g., bulk download) to CTS2 service framework ◦ Provide the capability to use data from one endpoint to query another ◦ Provide a framework that includes both CTS2 and Extensions ◦ Bulk downloads/special user requirements • Determine next steps/roadmap
2:00 PM - 4:00 PM	4W030	Information Session: Using LexEVS Query API and CTS2 API • Overview to demonstrate how to query using LexEVS APIs • Review of existing LexEVS API documentation • Open discussion - specific use cases

Attendees: Larry, Kim, Craig, Cory, Scott, Gilberto, Rob, Sherri, Joe (CTRP), Tin, Jason, Jose, Hemant (CTRP), Charlie

Discussion: Enhancing the LEXEVS CTS2 REST Interface

• CTRP Requirements

- Jose - would like to understand better ways for CTRP to utilize APIs - biomarkers (caDSR), indexing. In the process of rebuilding their application, and will want to look at considerations.
- Scott gave brief overview of LexEVS and support APIs - CTS2, Remove API, Java API
- Jose noted that CTRP uses the Java interface and REST interface. As they move forward, they would like to move to REST interface.
- Scott noted certain limitations - specifically limited search functionality currently implemented (note, this is limited by the implementation, not CTS2).
- Hermant - Discussed use cases
 - Identify the disease terms from NCIT - this is a process of abstraction to traverse hierarchy.
 - Scott agreed this could be accomplished with minor changes.
 - Complete term dump - SDC, ICD-9, ICD-10 - to be used to verify/validate codes.
 - Larry noted there is a bulk download available.
 - Need to be able to determine if things have changed after bulk download.
- Scott suggested that CTRP provide detailed use cases with input and output expectations.
- Jose discussed the use case that during a look up - to determine what it is a synonym of and what is the preferred term.
- Scott described searches - LexEVS Free text has 12 different way to search.
- Hermant noted they would want to know if something changed. Scott suggested that the versioning/history API may provide some of that information.
- Jose described there existing system. Currently it stores the terminal Leafs and parental terms
- Scott demonstrated the CTS2 API to get targetof/subjectof with the REST API
 - http://lexevs63cts2.nci.nih.gov/lexevscts2/codesystem/NCI_Thesaurus/version/15.08e/entity/C2991/targetof
 - For CTRP to get essentially perform transitive closure, they would need to call the server for each entity.
 - **Transitive closure would need to be supported for CTS2.**
 - This is specified in CTS2, but would need to be implemented in the LexEVS service
 - ancestors - A DirectoryURI that resolves to the transitive closure of the "parents" relationship(s). The primary purpose for this attribute is to provide a handle for subsumption queries. As an example, to determine whether Class X was a subclass of ClassY, one would query whether the EntityReference to Y was a member of X. ancestors.
 - <http://www.omg.org/cgi-bin/doc?formal/2015-04-06>
- Tracy shared that the path is currently stored in the database (transitivity table).
- Joe indicated that REST extensions over LexEVS make sense.
- Hermant described use case - the meaning of CCode has changed.
 - Gilberto noted that the history tables are published.
 - Gilberto discussed possible changes to a CCode - spelling error, hierarchy change, etc.
 - Larry noted the meaning of a concept doesn't change. If it changes, it is deprecated.
 - If you query against a deprecated concept, it will be returned and provide a pointer to the replacement concept.
 - Tracy noted that Sharon Gehene's group has an application that reconciles the changes.
 - Gilberto noted that we should **look at the History API and look at exposing the history API through CTS2.**
- Scott suggested that resolution of a large hierarchy is resource intensive. A graph DB implementation may provide this functionality.
- Jose indicated that their application needs to be complete by end of September 2016. **Any LexEVS related updates would need to be available well in advance of 2016 Sept to support CTRP development.**
- Gilberto demonstrated the use of the NCIT Browser to show how to build a graph (as described by Kim this morning).
- Tracy noticed that core:name doesn't seem correct for the core:entity
 - http://lexevs63cts2.nci.nih.gov/lexevscts2/codesystem/NCI_Thesaurus/version/15.08e/entity/C2991/targetof

- It was suggested that we **look at how this is implemented**.

Other issues:

- Federated Query of Resources
 - Gilberto described the use case to be able to query across different CTS2 services.
 - URI resolver won't necessarily solve this problem.
 - Larry - licensing may be a concern, especially with MedDRA.
 - Scott suggested this could be accomplished with existing code.
- Full resolution against value set, code system resolutions, queries
 - Scott described how to get the count of result by using HEAD in HTTP.
 - **The count should be verified** - as in the Java API as per estimates.
 - This should be **logged to CTS2** (<http://www.omg.org/issues/cts2-rtf.open.html>) as a request for -1 to indicate all.
- Compact URL to call to get particular value set results.
 - Not enough definition to respond.
- From earlier discussions
 - **Include ability to restrict to association**
 - **Include ability to provide complete transitive expression**

Information Session: Using LexEVS Query API and CTS2 API

- Scott demonstrated how to view a CTS2 query count by issuing HEAD instead of GET. This will provide a way to return all associated metadata.
- Scott reviewed LexEVS 6.x API documentation.
- Tracy noted that there are many Code snippet pages and we should **review these pages to determine if still valid or not**.
- Gilberto discussed the ability to pull in sample code to compile and run. There is nothing available now on the wiki.
- **"LexEVS Code Examples" to be reviewed and verified. (LexEVS Code Examples)**
- **Tracy is going to update the LexEVS Code Examples.**
- [LexEVS 6.x CTS2 API Quick Start](#) is documented better than the API.
 - **Include the call to HEAD**

Next Steps

- CTS2 Implementation Considerations:
 - Include ability to restrict to association
 - Include ability to provide complete transitive closure
 - Include ability to provide history API functionality
 - Review browser requirements to include all needed functionality
 - Include ability to group value sets together (FDA and SPL value sets)
- Documentation Considerations:
 - "LexEVS Code Examples" to be reviewed and verified.
 - Update CTS2 API documentation.

Wednesday, December 9, 2015

8:00 AM - 9:00 AM	2E914	Discussion: Review Lucene
--------------------------	-------	----------------------------------

Attendees: Kim, Cory, Scott, Craig

- Kim reviewed the work had has done with the Loader tutorial code that Scott provided.
- Scott discussed how property values are tied to entities in Lucene.
- It was noted that the 6.4 Lucene development isn't complete so some of the functionality is not working today.
- Scott reviewed how set theory is currently implemented. (2 parent block join)

9:00 AM - 9:30 AM	2E914	Recap and Planning
--------------------------	-------	---------------------------

Attendees: Kim, Cory, Scott, Craig, Jason, Larry, Tin

- The [2015.12 Technical Face-To-Face Prioritization List](#) was reviewed and updated to capture additional items.
- It was noted that **Partonomy should be considered for CTRP requirements**.
- Larry discussed structured presentation of Value Set and Mapping data. Currently flat lists (concepts with terms) or terms with source and target. To be useful, would **prefer to have hierarchy viewable to represent the internal structures**. There may be existing JIRA items, but we should look at this again. Scott suggested the use of codedNodeGraph call to create this hierarchy. Usage needs to be considered - requirements need to be established from the users and then look at the technology to support.

9:30AM - Noon	2E914	Discussion: Coding Scheme Search and Indexing - Determine requirements/use cases for horizontal coding scheme searches - Overview of indexing of qualifiers - Overview of search results in 6.4
		Discussion: LEXEVS Loader Improvements - Discuss known issues for OWL2, MedDRA and HL7 Loaders - Other loader improvements - Determine next steps/roadmap

Attendees: Kim, Cory, Scott, Craig, Jason, Larry, Tin, Rob, Tracy

Discussion: Coding Scheme Search and Indexing

- Traversing Coding Schemes
 - OBI and GO would be the starting place to determine ability needed to traverse from one coding scheme to the next. We have this captured and will consider.
- Indexing
 - Index Qualifiers
 - Scott described how we currently index qualifiers. Qualifiers are stored in a file as a list that are grouped together and parsed into the index. The list is added to the parent document as part of the block join implementation.
 - LexEVS 6.4 Implementation
 - Scott discussed the status of 6.4 and noted that we've noticed some result differences in going from a single index to multi indexes. The scoring is based on the frequency of a term - and we can boost the score. There is a junit that tests the boosting of terms, but we aren't sure this is a credible issue or not. Approximations are going to make this difficult. **Larry suggested that the raking be considered only for the individual source and not across all sources.**
 - Gilberto described a search result page where results could be split by vocabulary sources. To do this, the list of sources could be presented to the user and the user could select the source from a pop-up. The browser would need to be updated.
 - Even with multiple indexes, **exact matches will always be at the top of the results. Similar weighting should also be preserved.**
 - Larry requested that we **share with the group how it worked in 6.3 and how it now is returned in 6.4 once fully implemented.**
 - Scott noted that we could always write our own analyzer, but then we'd need to maintain and support.
 - Stop word list is still valid in new implementation.

Discussion: LEXEVS Loader Improvements

- OWL2 Loader
 - LEXEVS-586 - Getting issue details... STATUS
 - Used to support inferencing. GO and OBI may have this already.
 - Examples include - hasUncle, hasFather, etc.
 - Gilberto to provide examples and we can discuss during future meeting.**
 - http://www.w3.org/TR/owl2-new-features/#F8:_Property_Chain_Inclusion
 - LEXEVS-1160 - Getting issue details... STATUS
 - This is related to OWL2 and should be considered part of OWL2 changes.**
- MedDRA Loader
 - LEXEVS-339 - Getting issue details... STATUS
 - This can be closed since we have a MedDRA loader.**
 - LEXEVS-1169 - Getting issue details... STATUS
 - This does not impact the load.
 - Need to create JIRA for "Semantic Type" missing**
 - Need to provide comparison of MedDRA loads to understand what is missing. May decide to not pursue based on what is identified.**
- HL7 Loader
 - LEXEVS-584 - Getting issue details... STATUS
 - This has nothing of value and can be closed.**
 - LEXEVS-1037 - Getting issue details... STATUS
 - Scott noted that coding schemes should be loaded as separate coding schemes. Historically, we've loaded as a single coding scheme. There are varying views on how it should be loaded. There are approximately ~200 coding schemes and most have few concepts in them.
 - Namespaces need to be fixed - Scott asserted that there is no supported coding scheme (additional metadata)**
 - Gilberto described that on HL7 webpage, you can view the coding scheme and the branch/structure. It is desired to provide that structure.
 - We should look at what HL7 provides and determine what needs to be done.**

- Process Automation
 - Tracy described the sequential loading of content (GO, ChEBI, etc).
 - This manual loading happens monthly.
 - There are 5 manual steps.
 - **Propose a way to provide the loader a version and as a result it would build a manifest object.**
- Additional JIRA items

◦ type key summary assignee reporter priority status resolution created updated due

 Jira project doesn't exist or you don't have permission to view it.

[View these issues in Jira](#)

- **This list needs to be reviewed by the group to determine if there are additional loader considerations.** If there are JIRA items that are no longer needed, they should be closed.
- Scott brought up the issue around a failed load and table locking.
 - This issue is related to 5.5.
 - LexEVS should be aware so it can fix the problem.
 - One option would be to update 5.6 or MariaDB.
 - **JIRA item needs to be included.**
- **LEXEVS-234** - Getting issue details... STATUS
 - **This should be considered in project backlog.**
- **LEXEVS-347** - Getting issue details... STATUS
 - Tracy noted this would be difficult as the indicator varies based on the source.
 - Scott suggested an evaluation would need to be completed to understand the methods used to indicate deprecated concepts.
 - It was suggested to focus on specific coding schemes - instead of a general approach. Priority should be for OWL2.
 - **This should be considered in project backlog.**
- **LEXEVS-459** - Getting issue details... STATUS
 - Tracy - currently there is a script to load all the value sets (load value set definition).
 - Propose that the VS could be added to a directory and point the script at the directory to load the value sets.
 - This will be further discussed during the Value Set Editor discussion.
- **LEXEVS-535** - Getting issue details... STATUS
 - **This issue can be removed.**
 - CLAML loader is no longer used. All JIRA items around CLAML can be closed.
- **LEXEVS-464** - Getting issue details... STATUS
 - This is part of the older OWL loader.
 - This may have been resolved in OWL2.
 - The OWL to loader is no longer the focus.
 - No decisions made.
- **It was noted that we should do a complete review of the JIRA backlog (ON HOLD) issues. Most issues associated to OBO may be closed after review.**

1:00 PM - 1:30 PM	2E914	EVS Tools (LexEVS & EVS Focus) Meeting
		• Open topics

Attendees: EVS Group

- Reviewed [2015.12 Technical Face-To-Face Prioritization List](#) with the group.

1:30 PM - 4:00 PM	2E914	Tutorial/Discussion: Loader Implementation and Requirements
		• Overview of LexEVS Loader development process • Implement a simple LexEVS loader (hands-on) • Discussion of entry points into various loaders for purposes of updating • Discuss automating data loads and customization

Attendees: Tracy, Rob, Tin, Cory, Craig

- Scott presented the loader tutorial.

- Tracy wondered if an OWL2 to LexGrid Model spreadsheet (mapping) existed. This does not exist, but would be helpful. **It is suggested that this mapping be documented.**

Thursday, December 10, 2015

9:00 AM - 9:30 AM	6E030	Recap and Planning
--------------------------	-------	---------------------------

Attendees: Tin, Cory, Craig, Scott, Yeon, Jason, Kim, Larry

- Reviewed [2015.12 Technical Face-To-Face Prioritization List](#) with the group

9:30 AM - Noon	6E030	Discussion: Triple Store/Graph Database • NCI to provide requirements/use cases • Discuss plans based on Triple Store Capability Research and Analysis wiki pages • Demo of ArangoDB (https://www.arangodb.com/) • Determine next steps/roadmap
		Discussion: Cloud Considerations • NCI to provide requirements/use cases for supporting a cloud environment
		Discussion: Build and Deployment Process • DevOps and continuous integration • Changes for build, deployment and distribution ◦ Changes to support 6.4 ◦ Considerations for Triple Store/Graph DB ◦ Considerations for Cloud Infrastructure • Mayo's continuous integration server • Demo of Docker • Determine next steps/roadmap

Attendees: Tin, Cory, Craig, Scott, Yeon, Jason, Kim, Larry, Cuong, Jacob, Sara, Larry, Gilberto

Discussion: Triple Store/Graph Database

- Mayo has looked at the report from the SI group from CBIIT.
- Larry indicated that **SPARQL query is the most focus for the NCIT**. Also ability to **federate queries across SPARQL end points**. Would like have consistent results across LexEVS, SPARQL.
- Jason and Kim have been working on a project
- Gilberto - there are no use cases prepared. However, there are things that a terminology server cannot provide. Would like to have more integrated services.
 - For example, if researching Cancer and looking for gene data (how do I glue this information together). If both are in RDF, then can query using all with SPARQL.
 - Another example, is data elements - are there other data that exist that might be appropriate for my research. Users can start to explore ontologies for this data discovery.
 - **Federation of data from other SPARQL endpoints is the primary interest.**
- Larry suggested that Instead of LexEVS - Hierarchy and traversals might be better implemented in SPARQL.
- Gilberto -
 - **Federated queries - yes, that is primary focus.**
 - **SPARQL doesn't need to support reasoning - however, some minimal reasoning may be considered.**
 - **Performance isn't priority, but it can't be a bottleneck.**
 - **LexEVS/CTS2 doesn't need to tie to the triple-store (all would't be exposed through the triple store)**
- Kevin provided an overview of "what does a terminology database need to do?" and reviewed Key value store, document store (mongoDB, CouchDB), relational db and graph db usage to satisfy specific functionality required by a terminology.
 - KVS - Key-Value store; DS - Document Store; RDBMS - Relation Database; GDB - Graph Database

Datastore Feature	Datastore Type that Performs Well
Store a resource with an ID	KVS, DS, RDBMS, GDB
Find a resource by ID	KVS, DS, RDBMS, GDB
Find a resource by a set of properties	DS, RDBMS, GDB
Find all edges of a resource	GDB, RDBMS

Traverse a graph	GDB
Compute subgraphs	GDB
Perform set operations on subgraphs	GDB
Calculate paths	GDB

- Need to best look at your requirements and needs when choosing the solution.
 - Kevin looked at Neo4J, OrientDB, and others by performing benchmarks to determine how well these tools were improving.
 - Overall, Kevin found arangoDB to be best all around solution. It is a mix of document and graph solution.
 - Modeling is open for documents, graphs, and key value pairs
 - Allows for Joins
 - Provides graph functionality.
 - Gilberto - does arangoDB provide SPARQL endpoint plugin? Kevin indicated that arangoDB may not be supportive of SPARQL.
 - <http://stackoverflow.com/questions/34015945/sparql-interface-for-arangodb>
 - Demo of arangoDB
 - CTS2 JSON for parts of SNOMED loaded into arangoDB.
 - Benchmarks attempted
 - Neighborhood (Qualifier value) - LexEVS and CTS2 does this
 - returns in less than a second
 - Decendants (Qualifier value) - more difficult as maxDepth -1 (all)
 - returns in just over a second
 - typically done by building a table to traverse
 - Leaves (Event) (Return all the leaves)
 - Expensive to do in a DB
 - SNOMED Event branch - return all the leaves.
 - 7300 returned in less than 2 seconds.
 - Sub-Graphs (value set resolution related)
 - SNOMED root note - all Event branch with everything below, all observation branch and all of organism branch.
 - Return how many in each branch and then provide intersection of these branches and see what is returned.
 - returns in 3 seconds.
 - all - 354,000
 - event - 8500
 - obs - 855
 - organism - 34000
 - intersection - 1
 - Slightly slower results on OrientDB.
 - Graph neighbors - count only
 - How many nodes are in the graph - is difficult in LexEVS
 - extremely fast result.
 - JOINS from nodes to edges
 - Joining the edges to the entity.
 - returns relation, to and from
 - Shortest Path to Root
 - Returns vertices and edges
 - Gilberto - how much difference were there between the reviewed tools?
 - Kevin - OrientDB and ArangoDB are similar. Neo4J is the most mature of all, but didn't have same performance and was more of a pure graph database.
 - Tracy - to satisfy the need for SPARQL endpoint is Neo4J best?
 - Kevin - suggests that ArangoDB is not the way to go for SPARQL requirements.
 - Kevin's usecases for using ArangoDB is based on performance and ability to quickly meet requirements of users.
 - Larry - how could this be used in combination with LexEVS and other tools.
 - Kevin - the use of multiple stores/services is becoming more common to accomplish specific tasks.
 - NG (Kim and Jason) have been working on SPARQL endpoint for LexEVS
 - Doesn't have to go through database layer so it is faster.
 - Kim demoed some working code as part of the browser.
 - Trees and Hierarchy is faster.
 - Continuing to review and understand how SPARQL can apply to EVS tools.
 - Larry - how difficult will it be to deploy triple-store and graph DB in the NCI environment?
 - Sara - if part of build and deploy (aside from security concerns) then the tools support team can use. (for example struts, spring, etc).
 - This impacts the DBAs more than systems. It depends if the project teams need DBA support.
 - CBIIT managed hosting (supported by infrastructure teams) is currently how EVS is supported.
 - Scott indicated graph queries in LexEVS could be supported by triple stores. However, metadata cannot be supported in triple store. ArangoDB for example, could provide metadata on the edges.
 - **Design would need to be considered to provide a hybrid solution.** The LexEVS API would need to be transparent to the users. The API would need to wrap content from triple store and LexEVS DB.
- The **Mayo and NCI team needs to clarify the strengths and weaknesses of each and determine how to best address.**

Discussion: Cloud Considerations and Discussion: Build and Deployment Process

Attendees: Tin, Cory, Craig, Scott, Yeon, Jason, Kim, Larry, Cuong, Jacob, Sara, AJ, Larry, Gilberto

- Scott described considerations for cloud usage
 - Auto deploy
 - Auto Scale resources
 - Uptime
 - Sharable instances
- Kevin noted that technical is starting to provide ability to deliver what cloud promises. **Cloud is more about changing your development lifecycle than hardware.** Cloud is not server virtualization.
- Kevin demo'ed the use of Docker with LexEVS
 - Sara asked if it is possible to take a docker image and use it on different tiers - by passing in a variable to let the application know what to configure at that tier.
 - Kevin - this is the idea - and those variables can be stored in version control. This simplifies the process.
 - **Micro-architectures and micro-services is what is important today. LexEVS fits this model well.**
 - Attempting to document a LexEVS install is complex.
 - Docker Example has:
 - LexEVS
 - LexEVS-cts2
 - LexEVS-remote
 - mysql
 - uriresolver
 - **This docker container configures a complete LexEVS environment .**
 - Kevin described the use of a Nexus server. Similar to maven, docker images can be hosted on a private or public nexus repository. Nexus has expanded to include docker support - internal docker repositories.
 - Sarah - The tomcat, mysql and os come from public repositories.
 - Application versions can be specified or simply pull the latest from the docker server.
 - Sarah - **CBIIIT is not ready to support Docker and won't be available by March 2016.**
 - **The goal is to provide docker images for on premise (NCI) installation or install to external cloud services as required by NCI.**
 - **Sara - considerations for storing configuration files - we need to consider how passwords and other information**

1:00 PM - 3:00 PM	1W030	Discussion: Value Set Editor (Authoring) • NCI to provide requirements/use cases
		Discussion: The Future of IbGUI • Discuss future requirements • Review issues - JIRA and others • Develop a roadmap to address technical debt • Determine next steps/roadmap

Discussion: Build and Deployment Process and Discussion: Value Set Editor (Authoring)

Attendees: Cory, Craig, Scott, Jason, Kim, Larry, Gilberto, Rob

- Dev Ops Discussions carried over from earlier this morning.
 - Continuous Integration
 - Cory discussed the Continuous Integration Server usage and how it is used by the Mayo development team.
 - There was discussion about how to include CI server functionality to provide value to both the browser team and LexEVS team.
 - Currently Jenkins is unofficially supported at NCI, but they are supporting what is needed by the project teams.
 - Mayo does have Travis and Jenkins, but suggests the use of Jenkins.
 - Jason - there is interest and some value, but limited given few (one) developers. **It was suggested that it would be best to discuss with the Dev Ops group before setting up something else.**
- Value Set Editor
 - Requirements
 - **Ability to maintain collections and sub-collections of value sets**
 - https://nciterms.nci.nih.gov/ncitbrowser/ajax?action=create_src_vs_tree&nav_type=valuesets#

[Expand all](#) [Collapse all](#) [Check all](#) [Uncheck all](#)

- ☒ [CDISC: Clinical Data Interchange Standards Consortium Terminology](#)
 - ☐ [CDISC ADaM Terminology](#)
 - ☐ [CDISC ADaM Analysis Purpose Terminology](#)
 - ☐ [CDISC ADaM Analysis Reason Terminology](#)
 - ☐ [CDISC ADaM Date Imputation Flag Terminology](#)
 - ☐ [CDISC ADaM Derivation Type Terminology](#)
 - ☐ [CDISC ADaM Parameter Type Terminology](#)
 - ☐ [CDISC ADaM Time Imputation Flag Terminology](#)
 - ☐ [CDISC CDASH Terminology](#)
 - ☐ [CDISC CDASH Concomitant Medication Dose Form Terminology](#)
 - ☐ [CDISC CDASH Concomitant Medication Dose Units Terminology](#)
 - ☐ [CDISC CDASH Concomitant Medication Dosing Frequency per In](#)
 - ☐ [CDISC CDASH Concomitant Medication Route of Administration](#)
 - ☐ [CDISC CDASH ECG Original Units Terminology](#)
 - ☐ [CDISC CDASH Exposure Dose Form Terminology](#)
 - ☐ [CDISC CDASH Exposure Dosing Frequency per Interval Terminol](#)
 - ☐ [CDISC CDASH Exposure Route of Administration Terminology](#)

- **Ability to efficiently load the resolved value sets**
 - Tracy - Value Set authoring doesn't occur often. The value set resolution happens every time there is a new version of NCIT.
 - FDA and CDISC value sets are based on NCIT concepts.
 - Report writer templates are used for individual value sets.
 - **Further review of value set workflow (including resolution) is needed to determine requirements and proposed changes.**

Discussion: The Future of lbGUI

Attendees: Cory, Craig, Scott, Jason, Kim, Larry, Gilberto, Rob

- Current usage of lbGUI:
 - Mayo team uses the GUI to verify loads. (Development)
 - Tracy uses when the admin scripts aren't working. (Admin)
 - Rob uses to cleanup (Admin)
 - Gilberto uses to determine if data is loaded correctly. (Data)
- Scott noted that the representation of data in lbGUI isn't always correct. It is better to look at DB to determine how things are loaded.
- Technology needs to be updated
 - Usage is becoming unstable
 - There are several known bugs.
- **The focus should be on expanding the admin scripts and move away from lbGUI.**
- **Further review of admin workflow is needed to determine requirements and proposed admin script changes and additions.**
- **lbGUI will continue to be minimally supported as needed by the development team.**
- **Consider ability to provide viewing additional metadata in the GUI.**

3:00 PM - 5:00 PM	1W030	Debrief • Prioritize • Determine next steps/roadmap
--------------------------	-------	---

Debrief

Attendees: Cory, Craig, Scott, Jason, Kim, Larry, Gilberto

- Reviewed and prioritized [2015.12 Technical Face-To-Face Prioritization List](#) with the group