

2016.12 Technical Face-To-Face Meeting Notes

Document Information

Author: Craig Stancl, Scott Bauer, Cory Endle
Email: craig.stancl2@nih.gov, scott.bauer@nih.gov,
Team: LexEVS
Contract: 16X237
Client: NCI CBIIT
 National Institutes of Health
 US Department of Health and Human Services

Contents of this Page

- [Wednesday, November 30th, 2016](#)
 - [User Group Discussion](#)
 - [RESTful API Discussion](#)
 - [Triple store/RDF Discussion](#)
 - [EVS Project Group Discussion \(During regular call-in time\)](#)
 - [LexEVS Mapping Discussion](#)
 - [Lucene Discussion](#)
 - [Overflow/Additional Topics](#)
- [Thursday, December 1st, 2016](#)
 - [Value Set management and workflow](#)
 - [Value Set and Mapping Data with Hierarchical structure Discussion](#)
 - [NCI Systems Discussions](#)
 - [FHIR and terminology services \(CTS2\)](#)
 - [OWL Restrictions in LexGrid Model](#)
 - [Overflow/Additional Topics](#)
- [Friday, December 2nd, 2016](#)
 - [LexEVS Admin](#)
 - [Prioritization and Debrief](#)
 - [Prioritization and Debrief \(Continued if needed\)](#)

The **purpose of this document** is to document the technical face to face meeting details between the NCI and the LexEVS Team.

2016 November/December Face-to-Face Meeting Notes

Wednesday, November 30th, 2016

Time	Location	Topics	Participants
9:00 AM - 10:00 AM	4W030	User Group Discussion <ul style="list-style-type: none"> • Team to share how they are using LexEVS and additional usage requirements they may have. • What components of LexEVS do you currently use? <ul style="list-style-type: none"> ◦ LexEVS API ◦ LexEVS Remote API ◦ CTS2 RESTful services • Are there LexEVS services that you would like to use, but they don't meet your requirements? • Are there any other road blocks preventing you from using LexEVS? • What version of LexEVS are you currently using? 	CTRP, caDSR, GDC

Attendees: Larry, Jason, Kim, Craig, Scott, Cory, John, Liz, Sima, Rui, Natalia, Tracy, Sana, Tin, Gilberto

Discussion Points:

- caDSR Team represented by Sima, Natalia, Vikram
- caDSR Applications that use EVS
 - Sentinel - Alerts for concepts, job that does concept clean up (compares concepts)
 - Curation tooling - links to concepts, and search results.
 - CDE Browser - Concepts used from search results.
 - Semantic integration workbench - concepts used from search results
 - CDEs
 - Utilize the NCIT, NCI meta,
 - Look into NCIT - will use the concept to describe the CDE.
 - Organizing concepts to build CDE terminology.
 - CDEs are used for forms (permissible values on forms)

- Tooling hasn't changed or been replaced.
 - Currently use JARS and put in /lib
 - Using Remote API today.
 - Recently removed EJB
 - Need to consider architecture in the future.
 - MDR is planning to architect a solution moving forward.
 - Currently searches are restricted to preferred terms.
 - Building data element - definitional information, preferred name
 - Existing CDE - pull back preferred name.
- Current Tooling Issues
 - Need to have a data load completed to PROD.
 - Confirmed data load and ready once things move to production.
 - New LexEVS Jars will be included in next release.
- Remote API Architecture
 - Issues
 - Replacement of JARS
 - Serialization of objects.
- Proposed Architecture?
 - REST-ful API

Decision Points:

- Identify current JAVA API usage by caDSR
 - EVS team to provide feedback as to how to do things better.
 - EVS team to ensure that if REST-ful API is created, functionality to be prioritized.

Time	Location	Topics	Participants
10:00 AM - 11:00 AM	4W030	RESTful API Discussion <ul style="list-style-type: none"> • Discuss requirements for continued development of REST services. This will include both CTS2 and separate REST-ful API. • Browser: Discuss requirements for remote API and CTS2 REST-ful API. 	

Attendees: Larry, Jason, Kim, Craig, Scott, Cory, John, Liz, Sima, Rui, Natalia, Tracy, Sana, Tin, Gilberto, Jacob

Discussion Points:

- Browser use cases reviewed.
 - Additional cases
 - History - the browser currently uses.
 - Security - for Medra or other licensed vocabularies.
 - May be able to use CTS2 APIs, but may need to have separate REST-ful services for customized/specialized content.
 - Current browser wouldn't use REST services.
 - Unknown coverage for REST possibilities. So additional investigation required.
- Previous F2F considerations.
 - Custom Lucene may need to be provided for clinical trials.
 - Group Value Sets - may be useful.
 - Restrict to Properties - need to better understand this usecase.
 - History - suggested by the CTRP but has some requirements in scope of the Browser.
 - Graph and Association
 - No need from caDSR
- Additional requirements
 - Bulk Download
 - ability to download full or part of a complete terminology.
- Align rest calls to support Moonshot API services.
- Make others aware services are available.
- Tracy suggested to look at Data.gov
- Browser team not going to use REST at this time
- caDSR not going to use REST at this time.

Decision Points:

- Identify Moonshot Clinical Trial API services to be supported by REST services.
- Identify ways to promote REST services.
- Identify possibilities of participating in Data.gov

Time	Location	Topics	Participants
------	----------	--------	--------------

11:00 AM - 12:00 PM	4W030	Triple store/RDF Discussion <ul style="list-style-type: none"> • Discuss what triple stores would be used for in parallel and in conjunction with LexEVS • Text searching • NCBO SPARQL white paper and it's implications • Searching on Roles (Gilberto, Kim) 	
---------------------	-------	---	--

Attendees:

Discussion Points:

- There is a pilot ongoing to review triple stores
 - selected 3 triple stores (StarDog, Allegrograph (<http://allegrograph.com/>), Viritouso) and have been working for the past 6 months
 - evaluations covering restricting operations, loading data, performance testing, examine security (secured and anonymous access via proxies)
 - nature of queries haven't been as representative of what is needed for production.
 - more focus on real use queries.
 - operations - hosting model not supported by CBIIT - so no support. (ie, patching support not provided)
 - all can be queried with standard SPARQL
 - will still want REST services available to the end users.
- Transition to SPAQRL gives raw access to the data (unlike API)
 - However, you need to understand the data - and this could differ from endpoint to endpoint.
- What do TripleStores provide - that differ (better than) LexEVS
 - Representation of Hierarchy
 - Level of expressivity
- 3 use cases for triplestore evaluation:
 - Expressivity (reasoning support)
 - Linked open data
 - use vocabulary as a "glue" between different data repositories. Ability to "join" distributed repositories.
 - Closer integration of vocabulary and meta data (part of the MDR).
- Report writer now uses the triple store
 - Loading of Value sets take much less time
 - Access of Value Sets is better, but not hugely different.
- Need to determine where Triple Store is better and where LexEVS is better.

Decision Points:

- Investigate where Triple Store usage may augment LexEVS.

Time	Location	Topics	Participants
1:00 PM - 2:00 PM	5E030	EVS Project Group Discussion (During regular call-in time) <ul style="list-style-type: none"> • User/content priorities for value set, mapping, and other services. • Specialized search and other capabilities for complex chemical names and genetic names. • Current capabilities and browser implementation • Possible expert system extension 	EVS project meeting

Attendees: Larry, Kim, Craig, Scott, Cory, John, Liz, Gilberto, Lori, Tin, Teri, Sharon, Sana, Nick, Nels, Margaret, George, Brenda, Abigail, Erin, Joanne

Discussion Points:

- Value Sets
 - Would like LexEVS to support production of value sets with more rich structure. (to more efficiently assemble this deliverable)
 - 100K downloads from FTP site. Fewer users use the browser to download the value sets.
- Mapping
 - On a mapping page, you can download an excel or cvs file.
 - ie, chebi has mapping.
 - For GDC - ICD9 or 10 coding - to be able to use NCIT coding, there needs to be way to translate between ICD9 and NCIT codes. There currently isn't a good way to do that today.
 - ie ICD9 - Breast cancer - corresponds to ABC in NCIT
 - Determine how such a map could be published (browser and LexEVS)
- Other Services
 - IUPAC - there are 2 flavors to be considered - but can be managed.
 - HUGO - the slash and hyphens have been problematic, but have been mostly resolved in the NCI Browser.
 - Review of 4 identified searching issues (differences between results in LexEVS and Protege).
 - NGram tokenizers may provide solution if we implement an Expert System.

Decision Points:

- Identify additional mapping requirements from the EVS Project group.
- Investigate the use of the expert system solution to support specialized search capabilities for complex chemical and genetic names.
- Investigate the usage of NGrams in Lucene to support specialized search.

Time	Location	Topics	Participants
2:00 PM - 3:00 PM	5E030	<p>LexEVS Mapping Discussion</p> <p>Determine requirements and propose solution for mapping.</p> <ul style="list-style-type: none"> User requirement: One terminology to many terminologies mapping. Other topics: Current, conditional, external relationships 	

Attendees: Larry, Kim, Craig, Scott, Cory, John, Liz, Gilberto, Tracy

Discussion Points:

- Use case provided by external LexEVS user.
- one to many (one terminology to many) is currently not a priority.
- There may have been time when loading maps from UMLS, but not sure why it wasn't completed. Brian may have more information.
- Ability to capture synonymous (non-) - Query API, Loader. Would want a use case to specifically describe this mapping.
- Consider loading MRMAP and review how it is loaded.

Decision Points:

- Investigate the MRMAP load and determine why that work wasn't completed.

Time	Location	Topics	Participants
3:00 PM - 4:30 PM	5E030	<p>Lucene Discussion</p> <p>Propose additional features of Lucene to be used within LexEVS.</p> <ul style="list-style-type: none"> Discuss specialized search use cases. Possible Lucene enhancements for coding scheme categorizations, auto complete aids, Lucene services. 	

Attendees: Larry, Kim, Craig, Scott, Cory, John, Liz, Gilberto, Tracy

Discussion Points:

- Facets - ability to perform categorical search
 - Coding Scheme Types
 - Value Set Categories
- Auto complete
 - Interest for the browser and caDSR
 - concerns about the results being overwhelming
 - might be more useful if combined with facets - i.e. search cancers with facets of neoplasms
- Elastic Search
 - There are still many custom analyzers
 - If possible to make search more portable - across LexEVS and TripleStore - may want to look at this.
- Prefer to have a single interface - instead of having the user decide if they use LexEVS or TripleStore.
- SOLR vs Elastic Search
 - SOLR documents more flat
 - Elastic Search document more complex.
- Search down NeoPlasms and then stop at a certain level. There is no mechanism to capture that. Similar cases for drug searches.
- <http://www.immport.org/>
 - John demoed - this uses facets and auto complete (based on 3 chars or more and typing speed).

Decision Points:

- Investigate ability to use Lucene Facets and identify where it could be used.
- Investigate ability to design a usable auto complete and where it could be used.

Time	Location	Topics	Participants	Resources
4:30 PM - 5:00 PM	5E030	Overflow/Additional Topics		

Attendees:

Discussion Points:

Decision Points:

Thursday, December 1st, 2016

Time	Location	Topics	Participants
9:00 AM - 11:00 AM	3W030	Value Set management and workflow <ul style="list-style-type: none">• Discuss requirements for value set version management and workflow management and supporting technology.• Rob to give a demo of their current workflow and the scripts they use.• Discuss latest issues on PROD.	Rob, Tracy

Time	Location	Topics	Participants
11:00 AM - 12:00 PM	3W030	Value Set and Mapping Data with Hierarchical structure Discussion <ul style="list-style-type: none">• Determine requirements and propose options to hierarchical structure and mapping.• Discuss how VS could retain their multiple hierarchical structure that it came from.• Discuss what changes would be needed to CTS2 for this.	

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

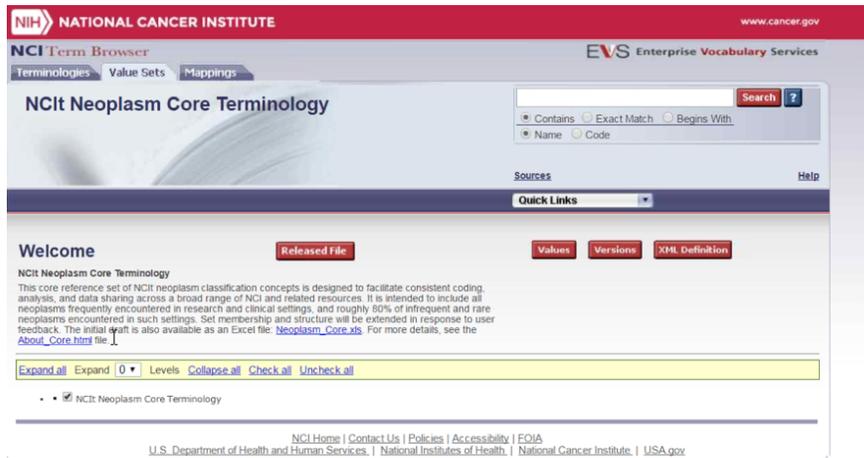
Discussion Points:

- Properties in Thesaurus that support the browsers
 - Subsets in Thesaurus (Protege)
 - Publish_Value_Set
 - Term_Browser_Value_Set_Description
 - Value_Set_Location - where browser fetches the report from (ftp location and path within evs, and filename, BNF)
 - TVS_Location - Terminology Value Set OWL file - hierarchy and components
 - Properties are scrubbed before loading into LexEVS - these are private/internal properties.
- Baseline - A diff is done on the value sets from month to month. Triggers update load procedures monthly.
 - Load, Remove and Resolve scripts are generated for changes
- OWL file provides information about where the value set lives in the hierarchy.
- NCI Thesaurus and TVS (provides structure to value sets) - used by the browser
- Process was created to provide structure/hierarchy to value sets
- Value Set loads - 700 coding schemes - loaded in 24 hours
- If value set resolution is performed against a new version of the code system, does LexEVS handle the versions?
- Process currently isn't limited to NCIT.
- Script created to create a txt file that views hierarchy of concepts on value set.
- value set downloads are ~10k a month and used across agencies by diverse set of consumers.
- curation and delivery processes are driven by the users and consumers.
- EVS editors work directly with CDISC and other groups.
- CDISC and FDA have different formats and standards
- Resolved value sets as coding schemes - done that way for performance.
- Process can be error prone for the EVS Editors. The editors need to do specific things to drive Browser.
- Hierarchy (value set groupings)
 - TVS_CDISC provides information for the browser to display the value set hierarchy

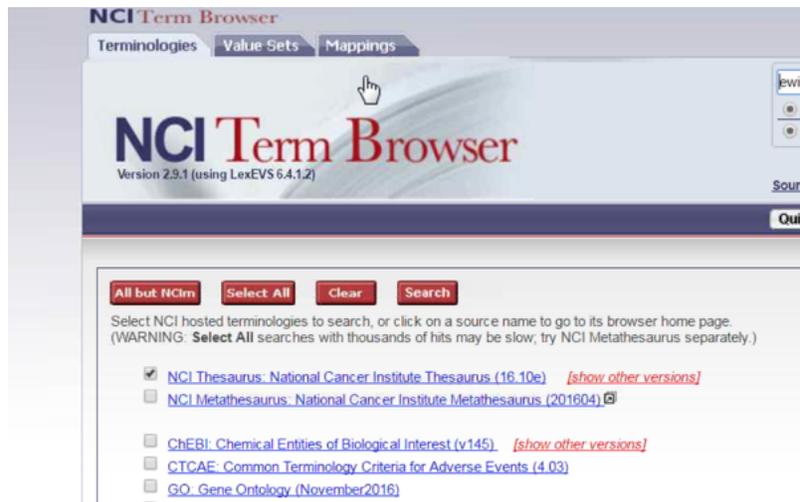


- Possibly add hierarch to the NCIT to replace the complexity today.
 - hierarchy (value set groupings) could be captured in lucene index (using Lucene Facets).

- Hierarchal representation in value set
 - A coding scheme with custom hierarchy



- Neoplasm core is a starting point for coding neoplasms. It is a starting point which then allows to branch out.
- Could extend the current implementation of resolved value sets so that that coding scheme would provide hierarchy. This is very much like a vocabulary.
- Could Provide "Hierarchal Value Sets"



- Browser could provide another tab "Hierarchical Value Sets" that would show the coding schemes that are the resolved hierarchical value sets.
 - This might require a different value set loader or an extension to it.
 - It could be complicated if the hierarchical value set hierarchy doesn't match the original coding scheme hierarchy.
- CTS2 representation would need to be extended to support the idea of a Hierarchical Value Set.
- Considerations for investigation
 - Investigate ability to be able to determine if Resolved VS coding scheme has changed.
 - Investigate ability to be able to determine if Value Set Definition has changed.
 - Investigate ability to update as needed (not have to load all 700 at the same time).
 - Investigate ability to capture value set groupings in Lucene index (using Lucene Facets and the NCIT).
 - Investigate ability to capture "Hierarchical Value Sets" as coding schemes with hierarchy.

Decision Points:

- Investigate ability to be able to determine if Resolved VS coding scheme has changed.
- Investigate ability to be able to determine if Value Set Definition has changed.
- Investigate ability to update as needed (not have to load all 700 at the same time).
- Investigate ability to capture value set groupings in Lucene index (using Lucene Facets and the NCIT).
- Investigate ability to capture "Hierarchical Value Sets" as coding schemes with hierarchy.

Time	Location	Topics	Participants
1:00 PM - 3:30 PM	3W030	NCI Systems Discussions <ul style="list-style-type: none"> • Nexus Deployment Discussion <ul style="list-style-type: none"> ○ Current status of LexEVS artifacts on NCI Nexus server ○ Discuss current technical challenges. • CI and Docker Status/Roadmap <ul style="list-style-type: none"> ○ Discuss the current status of the Docker scripts used to build/test LexEVS components. ○ Discuss NCI's current status and future plans to use Docker. ○ Discuss security challenges associated with NCI's environment and Docker. • Discuss a separate DEV environment for CI server deployment • Tech Stack Upgrades <ul style="list-style-type: none"> ○ Discuss DB upgrade: <ul style="list-style-type: none"> ▪ MySQL 5.6 vs. MariaDB (10.1 Supported 2017.01) ○ Discuss CentOS 7 upgrade • Tier Deployment testing responsibilities <ul style="list-style-type: none"> ○ Mayo development team responsibilities ○ NCI development team responsibilities 	Sara, Shireesha, Phil Jacob, Yeon (Systems Team)

Attendees: Larry, Sherri, Rob, Tracy, Jacob, Sarah, Scott, Cory, Craig, Kwan, Yeon, Sherri, Tin, Sana, Shireeseha, Jason

Discussion Points:

Nexus Server Configuration

- Testing on DEV tier - config of security permissions
- Manual publishing until configured
- CTS2 - Maven build was the simplest case, so that's the focus.
- ANT publishing not currently available. Will need to look at public and private key possibilities. Sara's team should be able to support. (LexEVS requires ANT build).

Tech Stack Updates

- DB
 - Currently at 5.5
 - Tech stack is moving to 5.6 and migration has started.
 - Preliminary tests show that we can support 5.6.
 - 5.6.33 is the current version. Yeon can update each tier for EVS team.
 - No plan for 5.7, but MarieDB (in 6 months to year)
 - Need to move to 5.6 as soon as we can.
- CentOS 7
 - LexEVS is ready to upgrade to CentOS 7
 - CentOS 7 is available.
 - Blade servers would need to be ordered or current blades would need to be re-imaged.
 - May be able to shuffle the upgrade and swap servers so it moves up.
- Java 1.8
 - LexEVS is ready with 1.8
 - Waiting on other tooling to support 1.8

Dev Environment

- Set up secondary Dev instance for Jenkins and application servers.
- Need to consider what database connection is needed.
- Suggested - set up a VM for this Dev
- Can submit tickets to Jacob to get this started.

Docker and CI Discussion

- Overview of Mayo usage and configuration.
 - NCI uses Jenkins 2.19
 - Docker differences between what NCI has and Mac version.
 - Would require move from Ubuntu.
 - Image repository not ready, testing Nexus 3.x for Docker Image Storage.
- NCI can support Docker configuration.
- Need to negotiate timelines.

Decision Points:

- Plan to migrate to 5.6.33 as soon as possible.
- Plan to migrate to CentOS7 (work with Jacob).
- Plan configuration of DEV instance (work with Jacob).
- Plan to further investigate Docker configuration.

Time	Location	Topics	Participants
3:30 PM - 4:00 PM	3W030	FHIR and terminology services (CTS2) <ul style="list-style-type: none">• Harold to provide update on CTS2 and FHIR.	Harold

Attendees: Tin, Jason, Rob, Tracy, Scott, Craig, Cory, Larry, Sherri, Harold, Gilberto

Discussion Points:

- Harold noted that the OMG process stalled by no further participation by Mayo.
- Remaining issues:
 - SOAP WSDL
 - Miscellaneous issues
- Additional Features:
 - Columnar Format
 - Canonical RDF
 - SNOMED CT implementation guide
- FHIR and CTS2 are similar in that both are complex but much is not required - only use what you need.
- Clinical Research and Biomedical informatics groups are taking note of FHIR and beginning participation in FHIR.
- FHIR Terminology - possible integration of CTS2 services.
- Grahm Grieve (HL7 FHIR) is in support of CTS2 services for FHIR.
- Current Planning
 - Plan on implementing entity description in native FHIR to demonstrate the differences and begin discussing with the FHIR community.
 - Review FHIR terminology and CTS2 terminology to describe overlap and gaps. A paper will be written and published.
- Other project - CIMI HSP - Determined that CTS2 wasn't a candidate for services.
- FHIR does offer:
 - Provides extensibility
 - Not to be fully implemented.
- HL7 and OMG HSSP process was not successful in that the standard wasn't successfully integrated back to HL7.
- A better model would have been what FHIR is doing within HL7 - collaborative within HL7.
- Harold to be at the January HL7 meeting to listen in on the FHIR sessions.

Decision Points:

Time	Location	Topics	Participants
4:00 PM - 5:00 PM	3W030	OWL Restrictions in LexGrid Model <ul style="list-style-type: none">• Discuss approach and propose additional features.• Determine if there are LexEVS model changes needed.• Loader considerations.• Additional problems and solutions	

Attendees: Larry, Sherri, Jason, Rob, Tracy, Cory, Scott, Craig, Harold, Gilberto, Kim

Discussion Points:

- Much of OWL2 is similar to OWL1.
- OWL2 includes property chains, but they aren't being used.
- The entire semantic meaning in LexEVS isn't required for OWL2. For example, reasoners would use OWL2 source - not out of a terminology server.
- There is no requirement to include additional OWL2 representation in LexEVS. Instead, use triple store and expand RESTful services.
- Current OWL2 issues have been resolved.
- Need to revisit the OBO JIRA item - and close it.

Decision Points:

- No additional OWL2 representation needed in LexEVS.
- Review OBO JIRA issue and resolve.

Time	Location	Topics	Participants
5:00 PM - 5:30 PM	3W030	Overflow/Additional Topics	

Attendees: Larry, Sherri, Jason, Rob, Tracy, Cory, Scott, Craig, Harold, Gilberto, Kim

Discussion Points:

- Browser issue - search issue when value sets don't return content.
- Noted that QA could be done in Protege before publishing
- Value Set Loader should not load a value set with no content.

Decision Points:

- Implement fix for Value Set Loader to not load value set with no content.

[LEXEVS-2510 - Getting issue details...](#)

STATUS

Friday, December 2nd, 2016

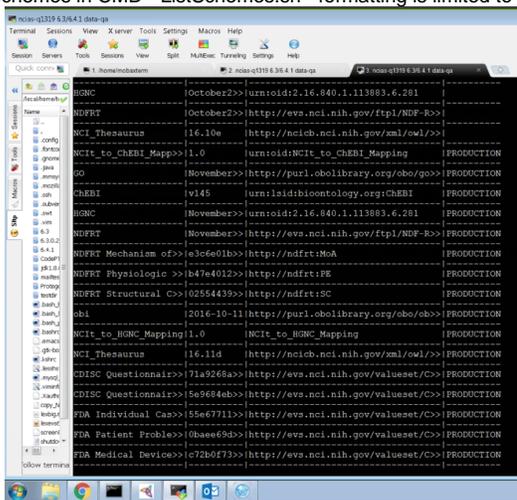
Time	Location	Topics	Participants
9:00 AM - 10:00 AM	1W030	LexEVS Admin <p>Discuss current and future requirements.</p> <ul style="list-style-type: none">• GUI<ul style="list-style-type: none">◦ Consider a web based tool. A simple way to look at the data.• Command Line loader requirements• Other considerations	

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

Discussion Points:

- Ability to look at data in a graphical way would be important.
- Command Line usage - List Schemes - can return 700+, so prefer to use the UI.

- Usage of the UI for troubleshooting to review the data in the database.
- Minimal ability to look at data would be preferred (fully graphical is not required)
- Administrative tasks not required in the GUI (only in the cmd line tooling)
- Ability to load metadata at the same time as the load.
- Web based tool.
- We need to replace based on functionality used by Browser
 - graphical hierarchy representation (tree extension?)
 - Optimal or best practice as an alternative
- providing code snippets for end users as options
- Secure any admin code (Loading, changing code systems) on a web based gui is a concern
- Could be potentially be used as a browser for technical users on an NCI Production server (Discussed)
- There is a request for admin ability for editing the preferences and manifest.
- Currently, there is no way to view what metadata is loaded.
- Investigate ability to combine data from the metadata and manifest files into one file.
 - This would make administration/loading easier.
 - Post load options may be an issue.
- History loader creates multiple errors when loading - there is an existing JIRA item.
 - May be caused by DB timeout.
 - Investigate what is causing.
- LG xml Loader is used to load maps. However, it doesn't take in account of the type of maps (it could). Not sure the rankings can be applied.
 - No existing issues, but may find some loading additional maps
 - SY relationships and Ranking are provided.
 - Monthly changes are applied.
- GUI Performance during x-forwarding noted by Rob.
- File system preferences - lock?
- UI is good for Tagging to Production
- UI is good for Removing a Coding Scheme
- Listing Schemes in CMD - ListSchemes.sh - formatting is limited to column width.



- Default, do not show entire width.
- Minimally add 10 chars to URL
- Minimally add 5 chars to Versions
- Add option to see full length's of all fields.
- Add option to see minimal information.

Decision Points:

Time	Location	Topics	Participants
10:00 AM - 12:00 PM	1W030	Prioritization and Debrief <ul style="list-style-type: none"> • Discuss OWL2, RRF, LexEVS, CTS2, Browser, and all previous topics <ul style="list-style-type: none"> ◦ Discuss future architecture • Determine next steps/road map and priorities 	

Attendees: Kumar, Larry, Jason, Sherri, Rob, Tracy, Cory, Scott, Craig

Discussion Points:

Architecture

- Future considerations.
 - Smaller services -
 - For example, Coding List listing service as a separate service.
 - Concerns around ability to deploy up the tiers
 - Current requirements will prohibit how quickly services can be exposed.
 - Concerns about tech stack upgrades across services. Micro Services may or may not be impacted by upgrades (some or all).
 - If addressed well, we can get rid of silos and duplication.
 - Resources are a concern,
 - Containers, JETTY, and how to balance.
 - Security
 - Scanning will take nearly as long as the large service.
 - Instead of re-architecting all, focus on new and additional functionality (along side existing LexEVS)
 - No longer would need clients to include jars, dependencies.
 -

Decision Points:

- Investigate services architecture to support new and additional functionality.

Time	Location	Topics	Participants
1:00 PM - 2:00 PM	1W030	Prioritization and Debrief (Continued if needed)	

Attendees:

Discussion Points:

Strategic direction - RESTful services

- Moving to micro architecture in new areas in functionality for LexEVS
- Integrated REST services across LexEVS, Triple Store, Clinical Trials (Integrated REST Services)
- Future MDR redesign effort - areas of service support of terminologies.
- Future CTRP support

Decision Points: