

**Observational Health Data
Sciences and Informatics (OHDSI)**

**Natural Language Processing (NLP)
Working Group**

11/4/2015

AGENDA

- ❖ Updates from the annual meeting
- ❖ Four projects
 - ❖ IRB for use of clinical text
 - ❖ Clinical text data storage and representation schema
 - ❖ NLP tools/pipelines for ETL
 - ❖ Use cases, e.g, phenotyping for cohort selection using NLP outputs
- ❖ Discussion

Project 1 - IRB for use of clinical text

- Based on Dr. Shah's (from Stanford) and Dr. Xu's (from UTHealth) clinical NLP IRB protocols, generalized protocol is being prepared
- Any additional protocols that can be shared would be welcome
- Next step – get approval from contributors and post it online

Project 2 - Clinical text data storage and representation schema

- Where to store text and derived data
 - Both original text and NLP outputs are stored outside of CDM
 - High-confidence NLP outputs go into the CDM tables, with a label of “from text”
- How to represent NLP outputs
 - SHARe/SemEval disease modifiers
 - cTAKES type system on medications and other entities?
 - Plan: combine all modifiers and create a minimum set of modifiers for all types of entities
- Next step
 - Presentation - Alex (Columbia) to work on the minimum set of modifiers?
 - Storage ?

Project 3 - NLP tools/pipelines for ETL

- Plan – develop a set of wrappers for multiple NLP tools, which convert each tool's outputs to OHDSI textual data schema
- Current status – cTAKES and MetaMap wrappers are being developed, focused on concepts only (not much on modifiers yet)
- Next Step/Discussion
 - A name for this tool
 - Two options of immediate actions: 1) positive concepts; 2) modifiers of concepts

Project 4 – Uses of NLP outputs - Phenotyping for cohort selection

- Plan – following Jon's system to build a keyword-based search engine for clinical text
- Current status:
 - Dataset: MTSamples 1000+ notes
 - Implementation: Elastic search
- Next Step/Discussion
 - A name
 - What to show – by patient, by note type (e.g., any note type terminology?)