



Scott L. DuVall

VA Salt Lake City Health Care System

University of Utah



@DuVallScott

pSCANNER: Lessons in Sharing

Data Sharing Plan

If an application describes a data-sharing plan,
NIH expects that plan to be enacted.

NIH Data Sharing Requirements



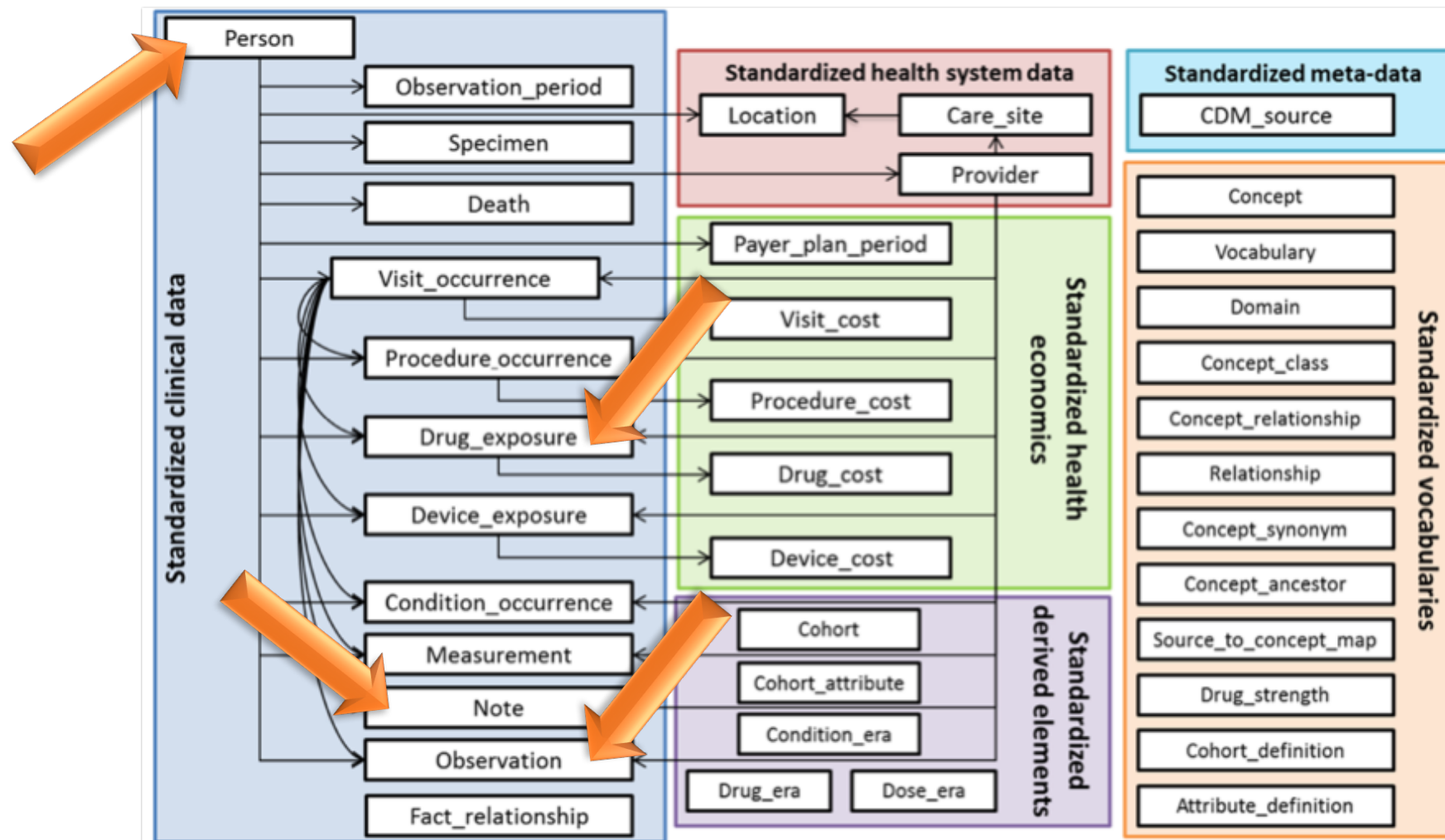
The diagram consists of three horizontal arrows pointing to the right, stacked vertically. The top arrow is dark blue and contains the word 'Data'. The middle arrow is light blue and contains the word 'Tools'. The bottom arrow is also light blue and contains the word 'Method'. The arrows are of increasing length from top to bottom, with 'Tools' starting further to the right than 'Data', and 'Method' starting further to the right than 'Tools'. All three arrows point towards the right edge of the frame.

Data

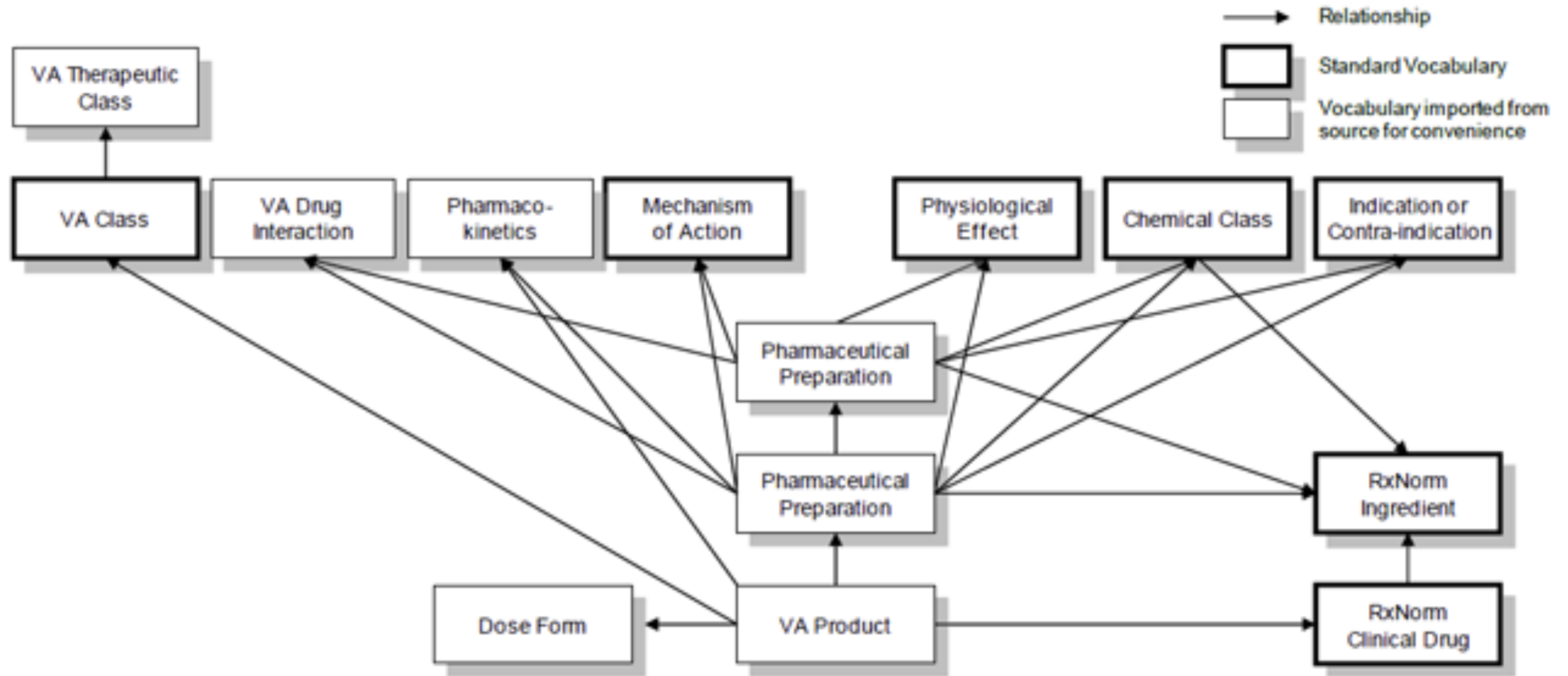
Tools

Method

Sharing Data



NLP Output in OMOP



NLP Output in OMOP



OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Note

Note_NLP

Note_Ontology

NLP Output in OMOP



The diagram consists of three horizontal arrows pointing to the right, stacked vertically and overlapping. The top arrow is light blue and contains the word 'Data'. The middle arrow is a darker blue and contains the word 'Tools'. The bottom arrow is light blue and contains the word 'Method'. The arrows are of equal length and are positioned such that they all point towards the right edge of the frame.

Data

Tools

Method

Sharing Tools

* Patterson OV, Freiberg MS, Brandt C, DuVall SL. Unlocking echocardiogram report measures for heart disease research through natural language processing. In prep.

Ejection Fr

ed at 35%

LVEF = 45-50

EF was 0.4



Sharing with CDRNs

Clamp Toolkit

+ New Project Run Tutorial A a

P Pipeline

- MyPipeline
 - clamp-ner
 - EfPipeline
 - Components
 - EfPipeline.pipeline
 - EfPipeline.pipeline.jar**
 - Named entity recognizer
 - POS tagger
 - Ruta rule engine
 - script
 - Section identifier
 - Sentence detector
 - Tokenizer
 - UMLS encoder
 - Data
 - PipelineLibrary

EfPipeline.pipeline

Move up Move down Delete Auto fix Edit Edit description

Name	Component	Description
DF_Clamp_sentence_detector	Sentence detector	Rule based sentence detector
DF_Clamp_tokenizer	Tokenizer	Rule based tokenizer
DF_OpenNLP_POS_tagger	POS tagger	OpenNLP based pos tagger
DF_Dictionary_based_section_identifier	Section identifier	Dictionary based section header Identifier
DF_CRF_based_named_entity_recognizer	Named entity recognizer	Name entity recognition using CRF
DF_Dictionary_based_UMLS_encoder	UMLS encoder	umls encoding algorithm

DESCRIPTION:
Simple pipeline to test CLAMP functionality

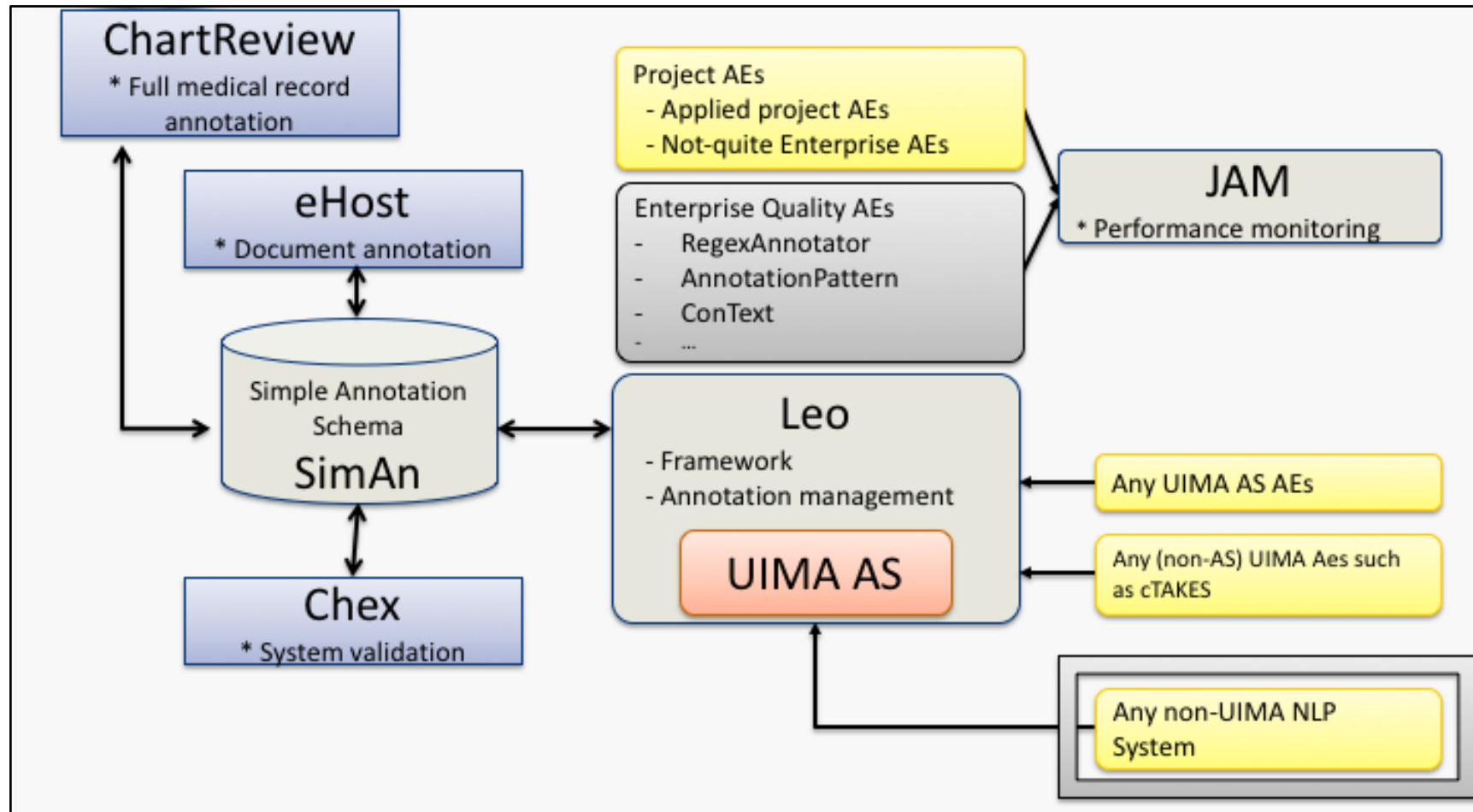
INPUT: All

OUTPUT: Name entities

CATEGORY:

71M of 237M

Sharing CLAMP



Sharing CLAMP

Clamp Toolkit

New Project

Run

Tutorial

A

a

P Pipeline

MyPipeline

clamp-ner

EfPipeline

Components

EfPipeline.pipeline

EfPipeline.pipeline.jar

Named entity recognizer

POS tagger

Ruta rule engine

script

Section identifier

Sentence detector

Tokenizer

UMLS encoder

Data

PipelineLibrary

EfPipeline.pipeline

Move up

Move down

Delete

Auto fix

Edit

Edit description

Name	Component	Description
DF_Clamp_sentence_detector	Sentence detector	Rule based sentence detector
DF_Clamp_tokenizer	Tokenizer	Rule based tokenizer
DF_OpenNLP_POS_tagger	POS tagger	OpenNLP based pos tagger
DF_Dictionary_based_section_identifier	Section identifier	Dictionary based section header Identifier
DF_CRF_based_named_entity_recognizer	Named entity recognizer	Name entity recognition using CRF
DF_Dictionary_based_UMLS_encoder	UMLS encoder	umls encoding algorithm

DESCRIPTION:

Simple pipeline to test CLAMP functionality

INPUT: All

OUTPUT: Name entities

CATEGORY:

71M of 237M

Sharing CLAMP

PAST MEDICAL HISTORY: 1. Stage III squamous cell lung cancer diagnosis in [**2534-2-21**], status post right pneumonectomy in [**2534-6-23**], with radiation, Carboplatin, and Taxol treatments. 2. Chronic obstructive pulmonary disease with PFTs in [**2535-5-23**] showing an FEV1 of 0.83 L, which is 25% of [** Location **]ted, and FEV1 to FVC ratio of 68% of [** Location **]ted. 3. Congestive heart failure with preserved left ventricular function in [**2535-5-23**]. 4. Atrial fibrillation. This was noted perioperatively. 5. Prostate carcinoma diagnosed in [**2531-2-21**] status post radical prostatectomy in [**2531-8-23**]. 6. Diabetes type 2. 7. History of urosepsis. 8. History of PE during the patient's postoperative course in [**2534-6-23**]. 9. Status post myocardial infarction. This was also perioperative in [**2534-6-23**]. Catheterization at that time showed normal left ventricular function, ejection fraction of 50%, and a 30% right coronary artery lesion. 10. Status post transient ischemic attack in [**2524**]. 11. Gout. 12. Gastroesophageal reflux disease. 13. Sleep apnea. 14. Colonic polyps noted in [**2532-5-22**]. 15. Hypercholesterolemia.

Annotation Types

☒ BaseT... ☒ Clamp... ☐ Docu... ☒ Segment ☒ Senten...

Click In Text to See Annotation Detail

- Annotations
 - ClampNameEntityUIMA
 - ClampNameEntityUIMA ("FEV1")
 - begin = 3098
 - end = 3102
 - semanticTag = test
 - assertion = null
 - cui = C0802965
 - attr1 = null
 - attr2 = null
 - attr3 = null
 - attr4 = null
 - attribute = {"umlsCuiDesc": "Volume^at 1.0 s post forced"
 - BaseToken
 - Sentence
 - Sentence ("which is 25% of [** Location **]ted, and FEV1 to")
 - begin = 3057
 - end = 3126
 - sentenceNumber = 72
 - segmentId = past_medical_history
 - Segment
 - Segment ("PAST MEDICAL HISTORY: 1. Stage III squamous")
 - begin = 2762
 - end = 3321
 - id = null

Sharing CLAMP

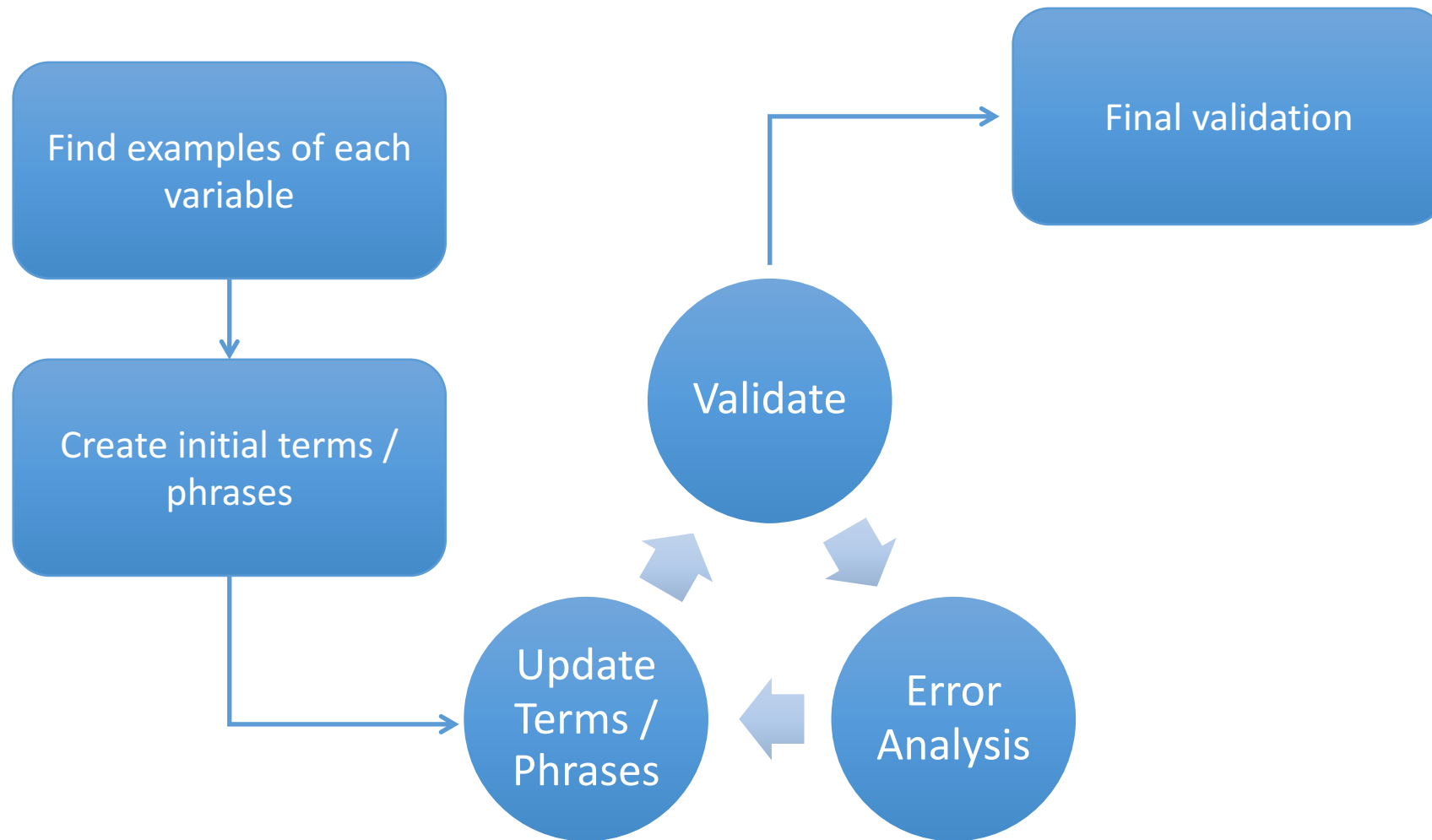


Data

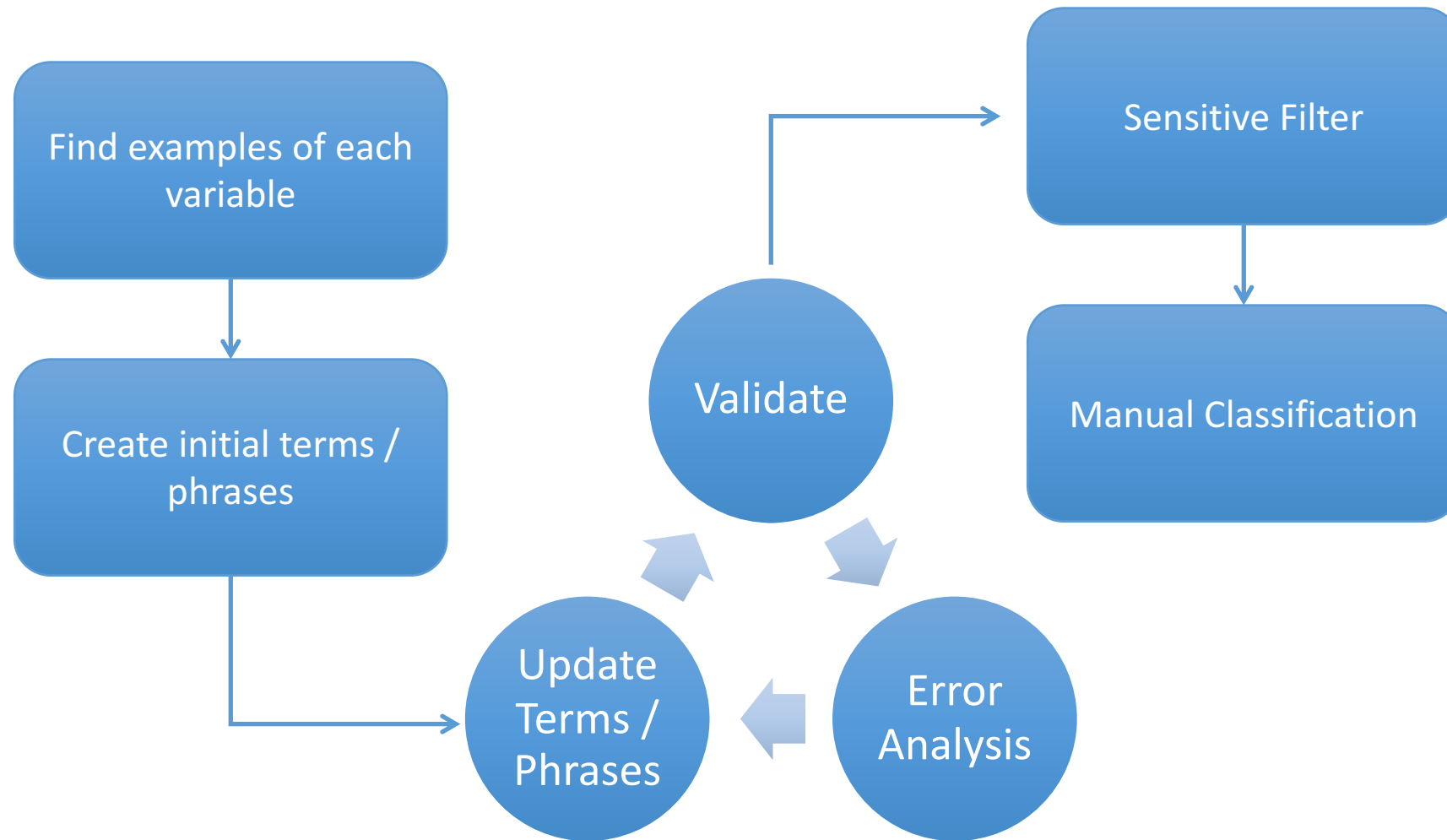
Tools

Method

Sharing Methods



Using NLP only



NLP-assisted Chart Review

Questions?

scott.duvall@va.gov



@DuVallScott