pSCANNER: Lessons in Sharing
If an application describes a data-sharing plan, NIH expects that plan to be enacted.
Sharing Data
NLP Output in OMOP
NLP Output in OMOP
NLP Output in OMOP
Sharing Tools
LVEF = 45-50%

Ejection Fraction was measured at 35%

Ejection Fraction was 0.4

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

CLAMP	Leo	Service
CLAMP	Leo	Service
CLAMP	Leo	Service
CLAMP	Leo	Service
CLAMP	Leo	Service
CLAMP	Leo	Service
CLAMP	Leo	Service

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

EfPipeline.pipeline

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

DESCRIPTION:
Simple pipeline to test CLAMP functionality
INPUT: All
OUTPUT: Name entities
CATEGORY:
PAST MEDICAL HISTORY: 1. Stage III squamous cell lung cancer diagnosed 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”]ed, and FEV1 to FVC ratio of 68% of [“Location”]ed. 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 [“2534”]. 11. Gout. 12. Gastroesophageal reflux disease. 13. Sleep apnea. 14. Colonic polyps noted in [“2532-5-22”]. 15. Hypercholesterolemia.
Sharing Methods

Data → Tools → Method
Using NLP only
NLP-assisted Chart Review

Find examples of each variable

Create initial terms / phrases

Validate

Sensitive Filter

Manual Classification

Update Terms / Phrases

Error Analysis
Questions?

scott.duvall@va.gov

@DuVallScott