

# User Friendly Natural Language Processing (NLP) of Free-text Clinical Notes at UCSF: EMERSE and cTAKES



William Brown III, PhD, DrPH, MA

Shivani Mehta, MPH



# Introductions



**William Brown III, PhD, DrPH, MA,** Assistant  
Professor of Medicine & Epidemiology and Biostatistics



**Shivani Mehta, MPH,**  
student in the Department of Epidemiology and  
Biostatistics

PhD

# CODE Lab

## Clinical and Observational Data Excellence

<code.ucsf.edu>

At CODE Lab our mission is to help researchers and healthcare workers contribute to creating a Learning Health System ... We accomplish these goals by:

- ▶ **Developing innovative methods and tools for collecting data**, improving adherence to biomedical protocols, and integrating diverse domain associated datasets.
- ▶ **Educating researchers, students, and healthcare providers** on the importance of informatics and on how to use informatics tools.
- ▶ **Harmonizing data for new analysis**, which will provide new knowledge to improve interventions and opportunities for new discoveries and collaborative research.

Lastly, **we committed to improving diversity, equity, and inclusion at UCSF** by leading existing efforts, and developing programs, that facilitate cross-cultural collaborations and recruitment and retention of underrepresented faculty, staff, and trainees.

The screenshot displays the CODE Lab website. The header includes the UCSF logo and navigation links: Home, People, Research, Tools, Media and Publications, and Commitment to Diversity. The main content area is divided into two columns: 'BASICS' and 'APPLIED'. The 'BASICS' column lists 'Biomedical Informatics Methods and Theories' and 'Human organizational factors'. The 'APPLIED' column lists 'Collecting and Curating Data', 'Individuals', and 'Populations and Cohorts'. Below these, a diagram shows the flow from 'Problem' (Small Subgroup, Needs, Controls, Small Sample Size) to 'Approach' (Ontology-based Data Integration) to 'Significance' (Increased Analytic Power). A 'Message from CODE Lab's Director' is also visible, featuring a photo of Dr. William Brown III and a quote about the lab's mission.

UCSF University of California San Francisco About UCSF Search UCSF UCSF Health

CODE Lab | Clinical and Observational Data Excellence

Home People Research Tools Media and Publications Commitment to Diversity

**BASICS**

BIOMEDICAL INFORMATICS METHODS AND THEORIES

Human organizational factors

Empirical discovery and prediction + Information management

**APPLIED**

COLLECTING AND CURATING DATA

Individuals

Populations and Cohorts

Computational biology

Clinical informatics

Public health informatics

Clinical research informatics + Translational bioinformatics

Biomedical Informatics Education

2 of 3

The Clinical and Observational Data Excellence (CODE) Lab was founded and is directed by Assistant Professor Dr. William Brown III. At CODE Lab our mission is to help researchers and healthcare workers contribute to creating a Learning Health System by leveraging New Media (i.e., mHealth, social media) and Biomedical Informatics (i.e., semantic harmonization, natural language processing, machine learning, data visualization) to collect data in real-time, improve treatment and adherence, and integrate gathered data among and between clinical systems and other research datasets.

**Problem**

Small Subgroup

Needs

Controls

Small Sample Size

**Approach**

Ontology-based Data Integration

**Significance**

Increased Analytic Power

Semantic Heterogeneity

Semantic Harmonization

Message from CODE Lab's Director

I started CODE Lab with the goal of helping to create a learning health system, where internal data and experience are systematically integrated with external evidence and that knowledge is put into practice. I believe that as a result, patients get higher quality, safer, more efficient care, and health care delivery organizations become better able to work. This is the

# GOALS of the Workshop

1. *Learn about NLP tools for research*
2. *Be introduced to user friendly NLP tools available at UCSF: EMERSE and cTAKES*
3. *Hear about "Use Cases" of NLP for Social Determinants of Health and Diabetes Research*
4. *Get help accessing these tools*
5. *Discuss and get your NLP questions answered*

# Agenda

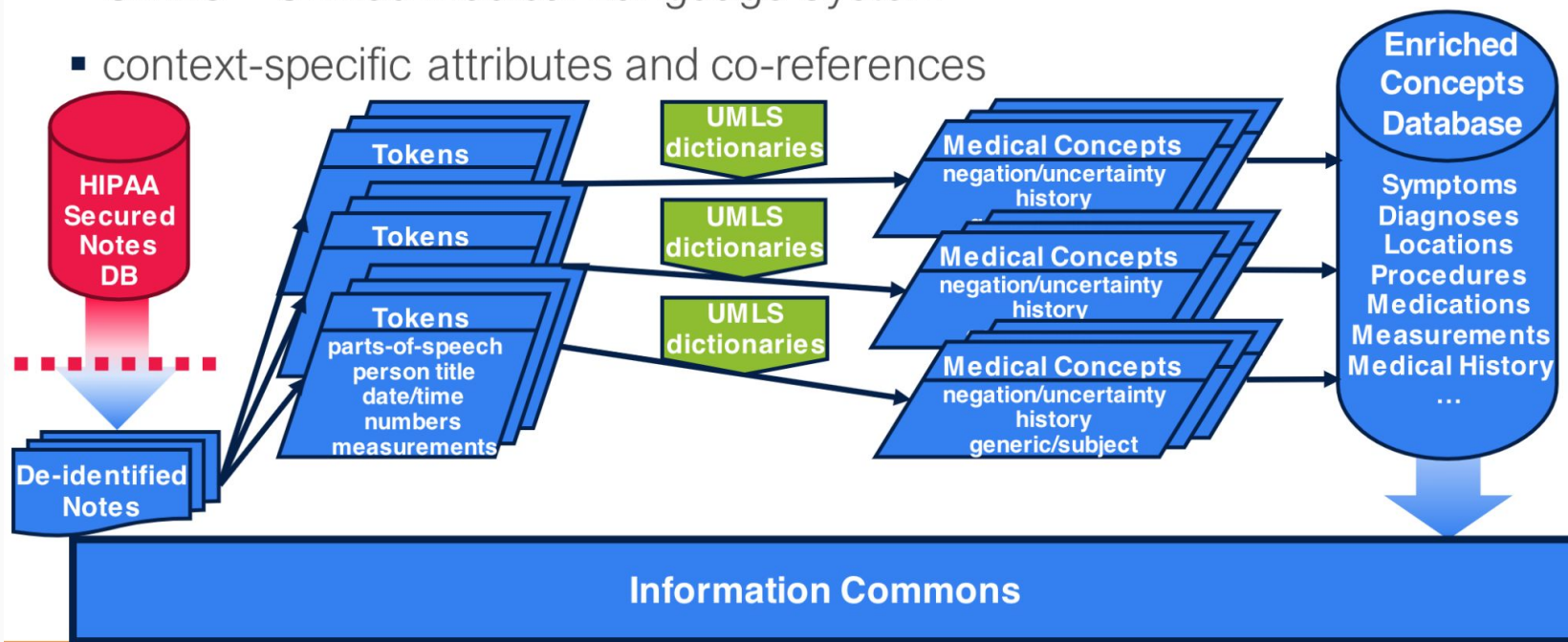
**Hour 1** (50 minutes) - Overview of NLP tools at UCSF  
(10-minute break)

**Hour 2** (50 minutes) "Use Case" presentation  
(10-minute break)

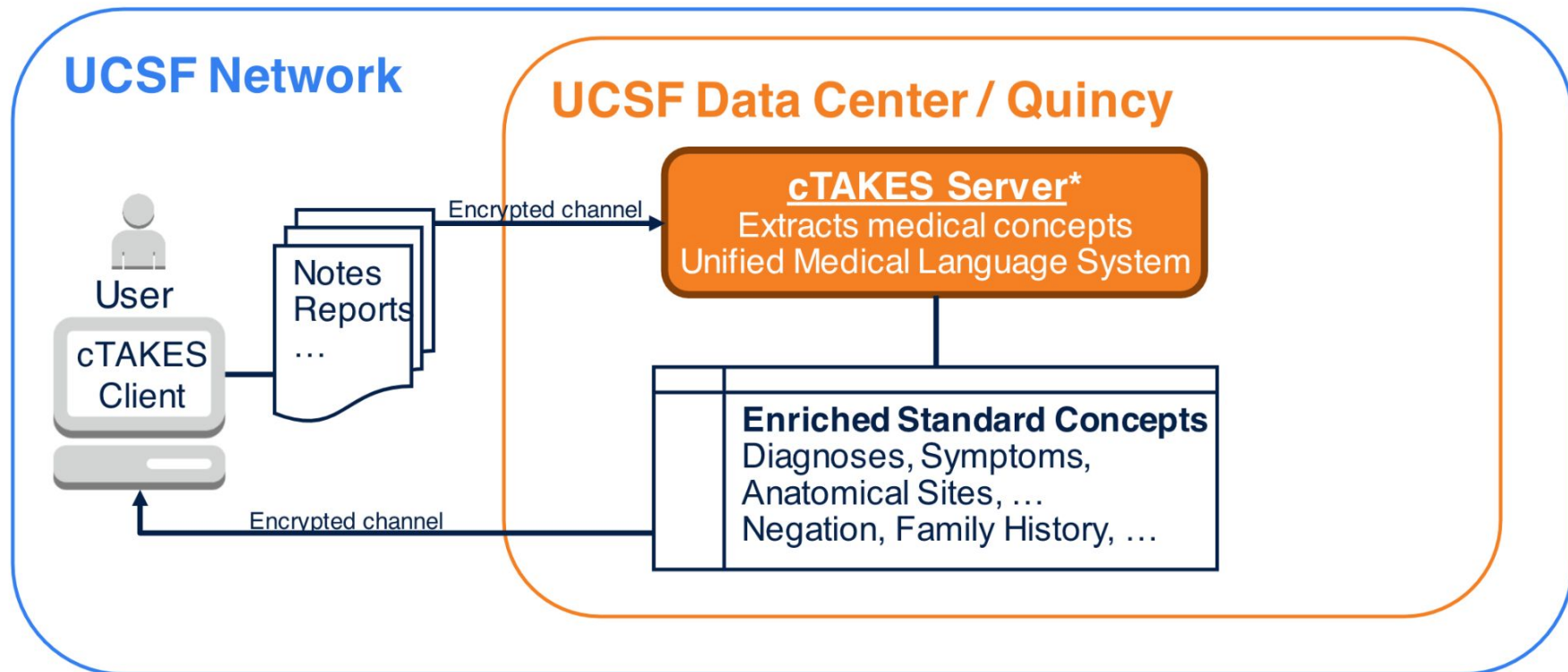
**Hour 3** - (50 minutes) Access Walk Through, Discussion, and Questions and Answers

## Information Extraction on Clinical Notes

- Apache cTAKES™ – clinical Text Analysis Knowledge Extraction System
- UMLS – Unified Medical Language System
- context-specific attributes and co-references



# cTAKES-as-a-Service PHI-safe User Environment



\*uses Apache cTAKES™ – clinical Text Analysis Knowledge Extraction System



# Concepts found by cTAKES

Ms. \_\_\_\_ also stated that she had just ended a three week course of **prednisone** four days ago, which she had started about a month ago at 60 mg and tapered herself down over a few days by 10 mg. She began the course of **prednisone** last month because she felt as if she was about to have a **Crohn's** flare at the time. Ms. \_\_\_\_ was last hospitalized at UNC for **Crohn's disease** exacerbation in March 2007. She denies any recent **hemoptysis**, **constipation**, **hematochezia**, **melen**, and changes in her **bowel habits** since Wednesday. She has been compliant with taking her medications for **Crohn's** and has been stable on her **mesalamine**, **mercaptopurine**, and **omega-3** acid supplement.

41: Fentanyl

Upon arrival to the ED, Ms. \_\_\_\_ was put on IV fluids, given **fentanyl** 50 mcg **IVP**, phenegran 12.5 mg diluted with 10 mL NS **IVP**, and Mg **sulfate** **IVP**. Radiological images were obtained through an abdominal CT **scan**, **ultrasound**, and 2V XR. Ms. \_\_\_\_ was not given any other narcotics for her **pain** because of a past violation of a **pain** contract after a positive toxicology **screen** for **cocaine** resulted in her discharge from her family medicine provider and due to suspicions that she was narcotic-seeking.

## Past Medical History

**Crohn's disease**, diagnosed 1998  
**Adenocarcinoma of terminal ileum** 1998 - s/p  
**resection of terminal ileum**, **rad** and **chemo**, no mets.

hx of **small bowel obstruction** secondary to **Crohn's Disease**  
 DM  
**HTN**  
**DVT** and PE 2001  
**PUD**

text	negated	location	domain	hx	conditional	cui	location snomed id	vocab term	vocab
<a href="#">Crohn Disease</a>	✗		<b>diseases</b>	✗	✗	C0010346		34000006	SNOM...
<a href="#">Crohn Disease</a>	✗		<b>diseases</b>	✗	✗	C0010346		34000006	SNOM...
<a href="#">Hemoptysis</a>	✓		<b>symptoms</b>	✗	✗	C0019079		66857006	SNOM...
<a href="#">Constipation</a>	✓		<b>symptoms</b>	✗	✗	C0009806		14760008	SNOM...
<a href="#">Hematochezia</a>	✗		<b>symptoms</b>	✗	✗	C0018932		405729...	SNOM...
<a href="#">Melena</a>	✓		<b>symptoms</b>	✗	✗	C0025222		2901004	SNOM...
<a href="#">Defecation</a>	✗	Intestines	<b>symptoms</b>	✗	✗	C0011135	261093006	39211005	SNOM...
<a href="#">Crohn Disease</a>	✗		<b>diseases</b>	✗	✗	C0010346		34000006	SNOM...
<a href="#">mesalamine</a>	✗		<b>medications</b>	✗	✗	C0127615		52582	RXNO...
<a href="#">mercaptopurine</a>	✗		<b>medications</b>	✗	✗	C0000618		1039008	SNOM...
<a href="#">Omega-3 Fatty Acids</a>	✗		<b>medications</b>	✗	✗	C0015689		4301	RXNO...
<a href="#">Fentanyl</a>	✗		<b>medications</b>	✗	✗	C0015846		373492...	SNOM...
<a href="#">Intravenous pyelogram</a>	✗		<b>procedures</b>	✗	✗	C0203108		32265006	SNOM...
<a href="#">Intravenous pyelogram</a>	✗		<b>labs</b>	✗	✗	C0203108		32265006	SNOM...
<a href="#">Intravenous pyelogram</a>	✗		<b>procedures</b>	✗	✗	C0203108		32265006	SNOM...
<a href="#">Intravenous pyelogram</a>	✗		<b>labs</b>	✗	✗	C0203108		32265006	SNOM...
<a href="#">Sulfates, Inorganic</a>	✗		<b>medications</b>	✗	✗	C0038720		95975008	SNOM...
<a href="#">Sulfate measurement</a>	✗		<b>procedures</b>	✗	✗	C0474702		104948...	SNOM...

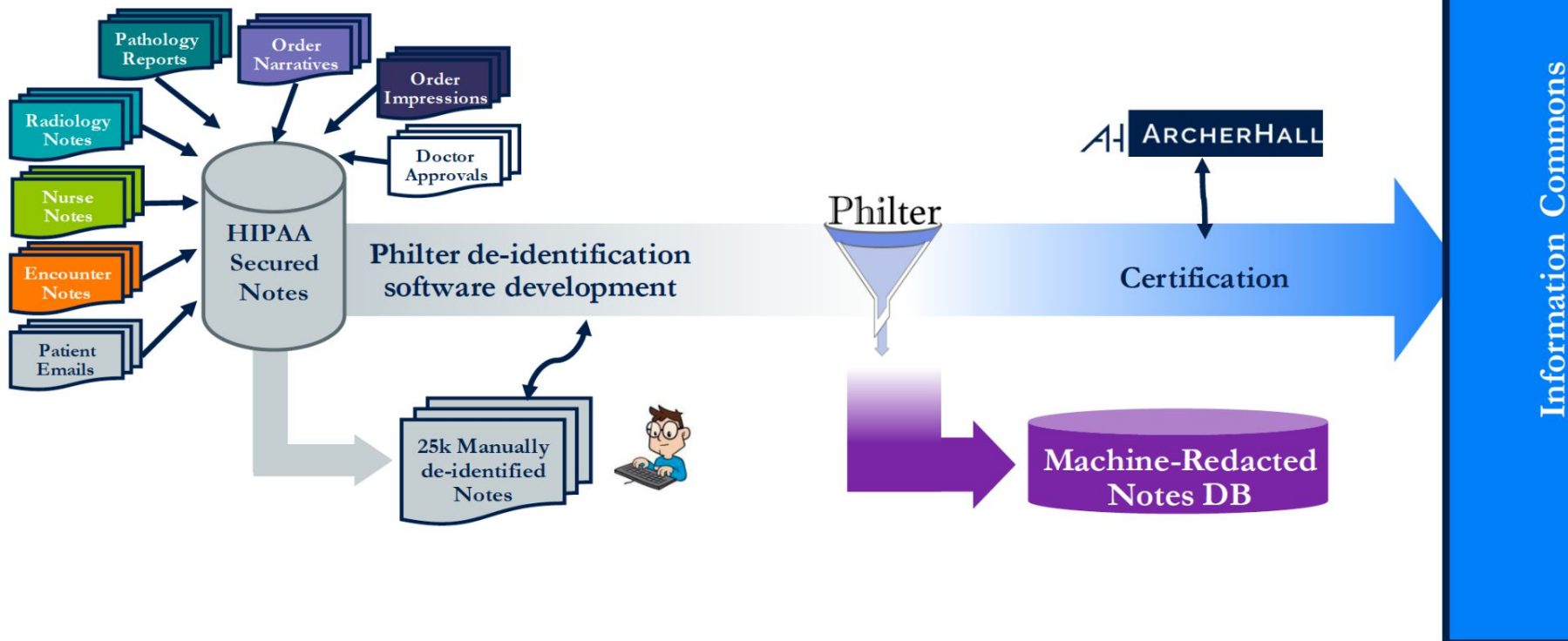


# EMERSE: Electronic Medical Record Search Engine

- Open-source software tool to search and browse notes data, without programming experience.
  - EMERSE was created by the University of Michigan.
- Clinical notes extracted from UCSF Electronic Health Record, Epic, and processed by software to remove 18 PHI elements.
- Patient's records are assigned synthetic identifiers that maintain the link between notes and the patient's structured data in DEID CDW.

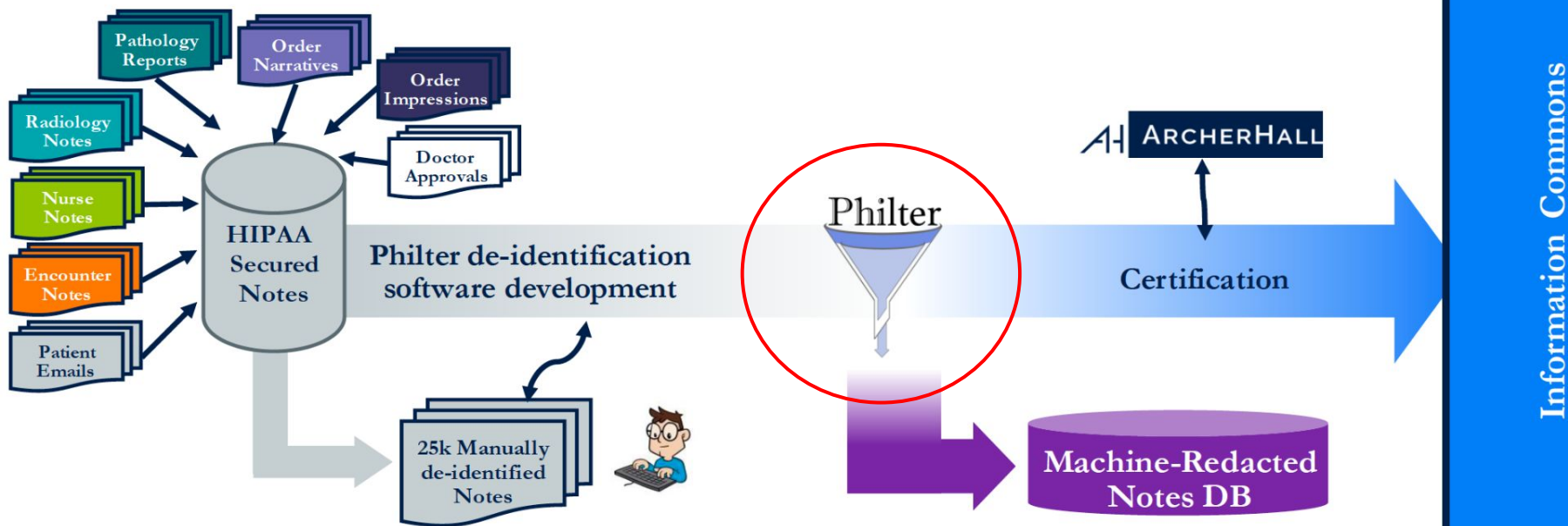
# Clinical Data and Notes

102 million clinical notes



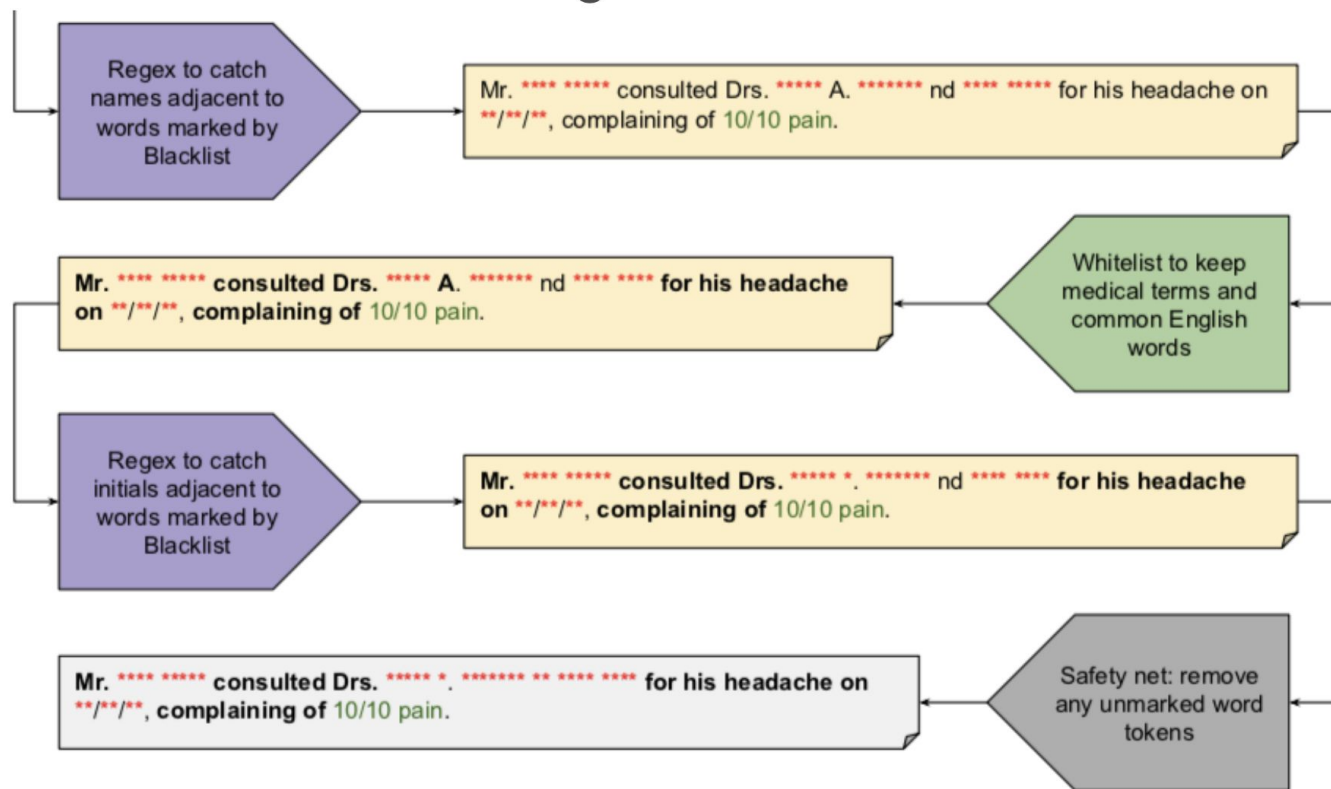
# Protected Health Information filter (Philter)

102 million clinical notes



Protected Health Information filter (Philter): accurately and securely de-identifying free-text clinical notes. NPJ Digit Med. 2020; 3:57. Norgeot B, Muenzen K, [Peterson TA](#), Fan X, Glicksberg BS, Schenk G, Rutenberg E, Oskotsky B, [Sirota M](#), [Yazdany J](#), [Schmajuk G](#), Ludwig D, Goldstein T, **Butte AJ**. PMID: 32337372; PMCID: [PMC7156708](#).

## Illustration of Philter Algorithm



Protected Health Information filter (Philter): accurately and securely de-identifying free-text clinical notes. NPJ Digit Med. 2020; 3:57. Norgeot B, Muenzen K, [Peterson TA](#), Fan X, Glicksberg BS, Schenk G, Rutenberg E, Oskotsky B, [Sirota M](#), [Yazdany J](#), [Schmajuk G](#), Ludwig D, Goldstein T, **Butte AJ**. PMID: 32337372; PMCID: [PMC7156708](#).

# Sample Note

## Before De-identification

A900001 2011-1-20 18:40:00.000 7465789 1

EMERGENCY DEPARTMENT CHART [FINAL] -----

- VISIT DATA -----

----- Patient SMITH, MARY (317 62 10-1), visit number 189487645, arrived at the ED at 1/20/2011 5:35:18 PM with

a complaint at registration of PAIN ARM/LEG-EXTREMITIES. ----- HISTORY OF PRESENT ILLNESS ----- (Triage complaint: brick fell on RT foot yesterday at work. now c/o rt big toe pain. no swelling UC ), 29 y.o. M c/o of R big toe pain 1/10 mild after a brick fell on his R foot yesterday. He denies any other injury or pain or complaint. No fall. Able to walk without pain or difficulty. No pain or complaint proximally up his RLE. No numbness/tingling., --Medications None-- Allergies None --Past Medical History rt eye surgery -- Tobacco: None --Alcohol: None --Recreational drugs: denies Other medical charts and records reviewed and interpreted: Triage and nursing notes ----- REVIEW OF SYSTEMS -----

----- --GENERAL ROS comments / Constitutional: negative, fever, chills. --EYES: negative . --EAR NOSE MOUTH THROAT: see HPI

## After de-identification

DB35CD2D4DD54A 01/05/2011 18:40:00.\*\*\*\*\* 1

EMERGENCY DEPARTMENT CHART [FINAL] -----

VISIT DATA -----

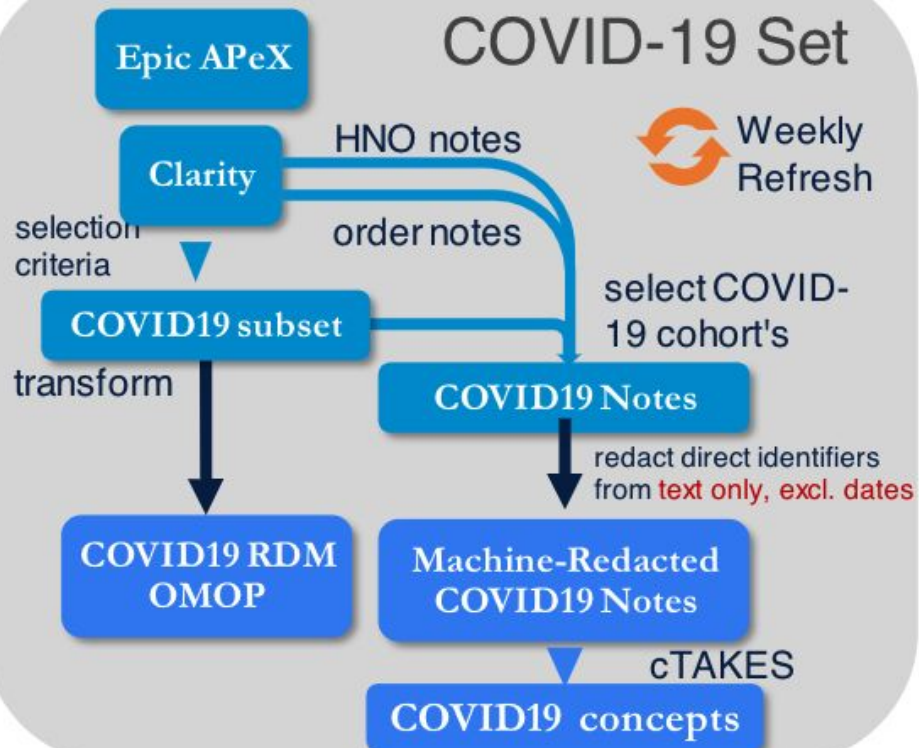
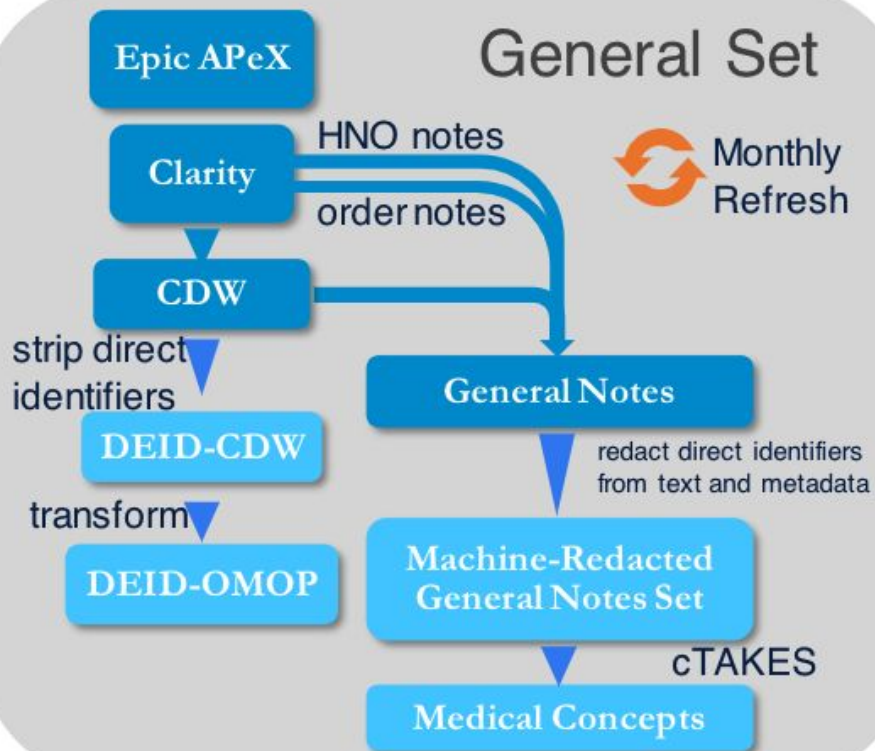
Patient \*\*\*\*\*, \*\*\*\*\* (\*\*\*\*\* \*\*\*\*\* \*\*\*\*\*\_\*\*\*\*\*), visit number \*\*\*\*\*, arrived at the ED at 01/05/2011 5:35:18 PM with a

complaint at registration of PAIN ARM/LEG-EXTREMITIES. ----- HISTORY OF PRESENT ILLNESS ----- (Triage complaint: \*\*\*\*\* fell on RT foot yesterday at work. now c/o rt big toe pain. no swelling UC ), 29 y.o. M c/o of R big toe pain 1/10 mild after a \*\*\*\*\* fell on his R foot yesterday. He denies any other injury or pain or complaint. No \*\*\*\*\*. Able to walk without pain or difficulty. No pain or complaint proximally up his RLE. No numbness/tingling., --Medications None--Allergies None --Past Medical History rt eye surgery --Tobacco: None --Alcohol: None --Recreational drugs: denies Other medical charts and records reviewed and interpreted: Triage and nursing notes ----- REVIEW OF SYSTEMS -----

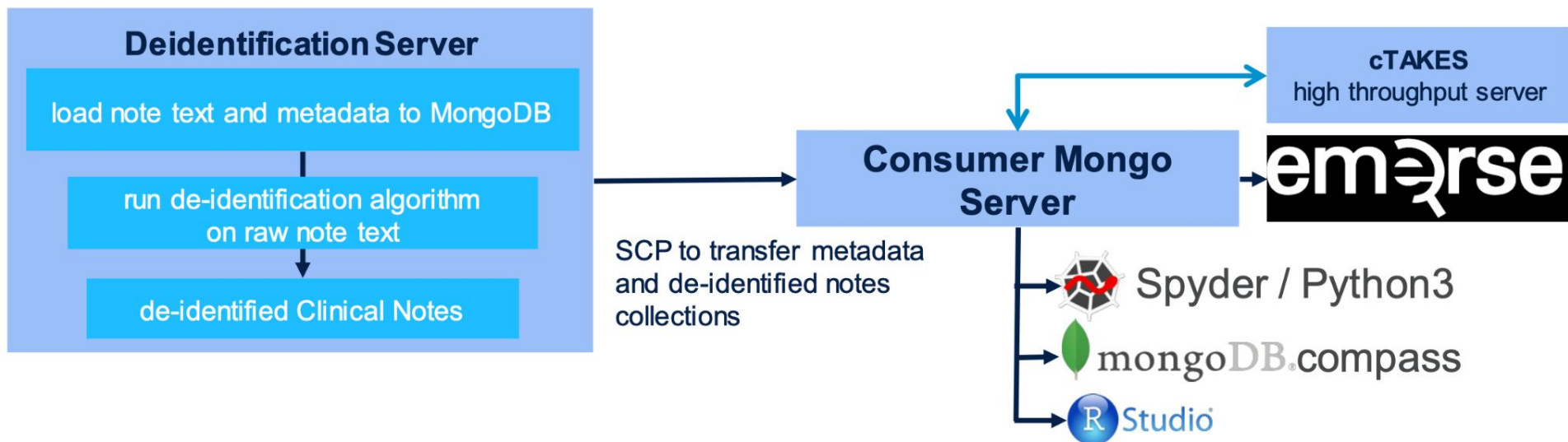
GENERAL ROS comments / Constitutional: negative, fever, chills. --EYES: negative . --EAR NOSE MOUTH THROAT: see HPI.




# UCSF Clinical Notes



# M-R Notes Data Flow





 EMERSE

William Brown ^

PatientsAll Local Patients (2,516,714)

Filters

Terms

social

amiable

friendly

"fun loving"

genial

gregarious

outgoing

sociable

Results

HIGHLIGHT DOCUMENTS

FIND PATIENTS

Temporary TermsSaved TermsAdvanced TermsSynonym Preferences

Temporary Terms

Name/Description

Manage Terms

View Terms

Sharing


Clear/Delete

Export

Terms to include


Please add terms or multi-word phrases, one at a time using the field below.


Add term


 Ignored


ADD


Included terms


 social


 amiable


 friendly

 "fun loving"

 genial

 gregarious

 outgoing

 sociable

UPLOAD TERMS

SAVE TERMS

Phrases to exclude

Please add phrases to exclude one at a time using the field below.

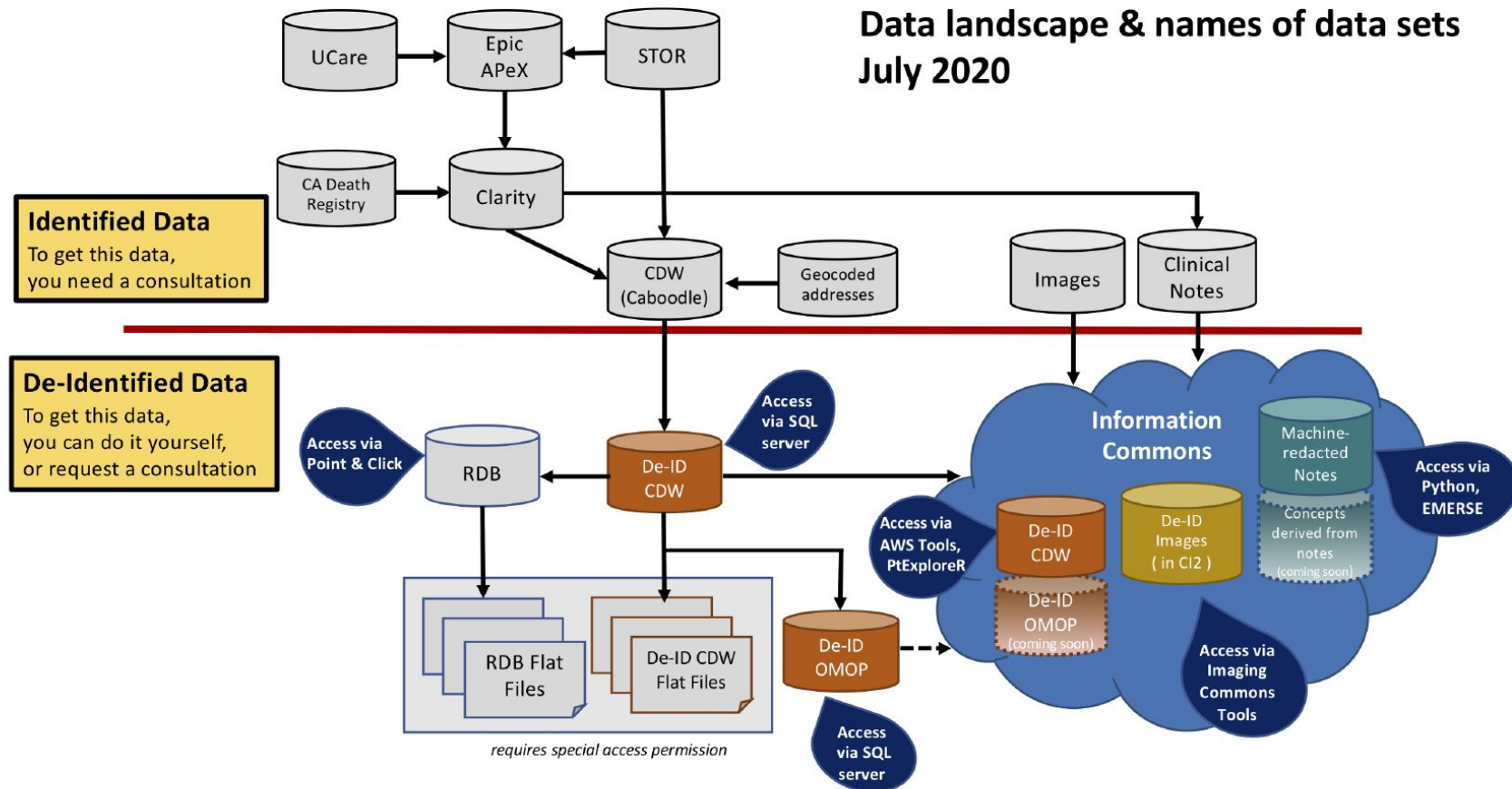
Add Phrase

ADD

Excluded phrases

None

## Data landscape & names of data sets July 2020



# Patient ExploreR

- User friendly interactive interface with drill-downs
  - <https://patientexplorer-i.c.ucsf.edu/app/pe>
- De-identified row level EHR data
- Search and export patient cohorts

UCSF PatientExploreR

Patient Finder Patient Explorer About Help Logout

Patient Finder

Search Mode: **SEARCH BY CRITERIA** ALL PATIENTS PATIENT LOOKUP

Patients in the system: 3,746,588  
Latest data refresh: 11/30/2021

Step 1. Filter by Search Criteria:

OR AND

+ Add clinical event + Add patient detail + Add group of items

Step 2 (optional). Filter by Event Sequences:

Select up to two ⓘ items in Step 1 to filter them by event

Simplified Search Preview:

View actual SQL

FIND MATCHING PATIENTS RESET SEARCH

**Event Selection Menu:**

- Clinical Event - Diagnosis
- Clinical Event - Lab
- Clinical Event - Medication
- Clinical Event - Procedure
- Clinical Event - Provider
- Clinical Event - Vitals
- Clinical Encounter

UCSF PatientExploreR

Patient Finder

[Patient Explorer](#)[About](#)[Help](#)[Logout](#)

Patient Finder

Patients in the system: 3,746,588  
Latest data refresh: 11/30/2021

Search Mode:

SEARCH BY CRITERIA

ALL PATIENTS

PATIENT LOOKUP

Step 1. Filter by Search Criteria:

OR

AND

+ Add clinical event

+ Add patient detail

+ Add group of items

Clinical Event - Diagnosis

Clinical Event - Lab

Clinical Event - Medication

Clinical Event - Procedure

Clinical Event - Provider

Clinical Event - Vitals

Clinical Encounter

Step 2 (optional). Filter by Event Sequences:

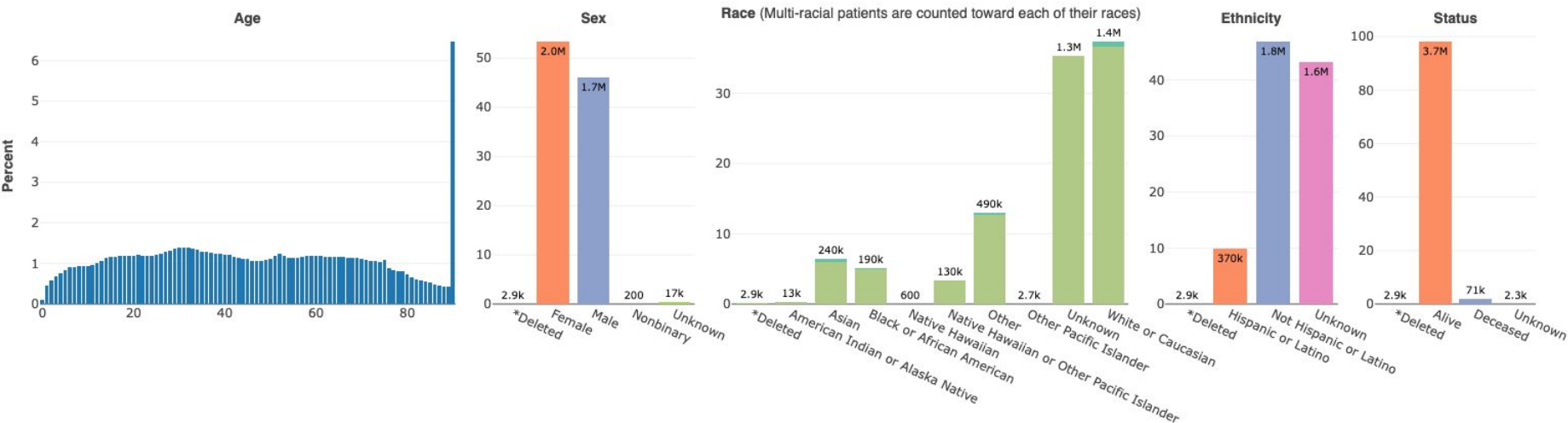
Select up to two ⓘ items in Step 1 to filter them by event sequence

Simplified Search Preview:

FIND MATCHING PATIENTS

RESET SEARCH

# Patient ExploreR



# Questions?



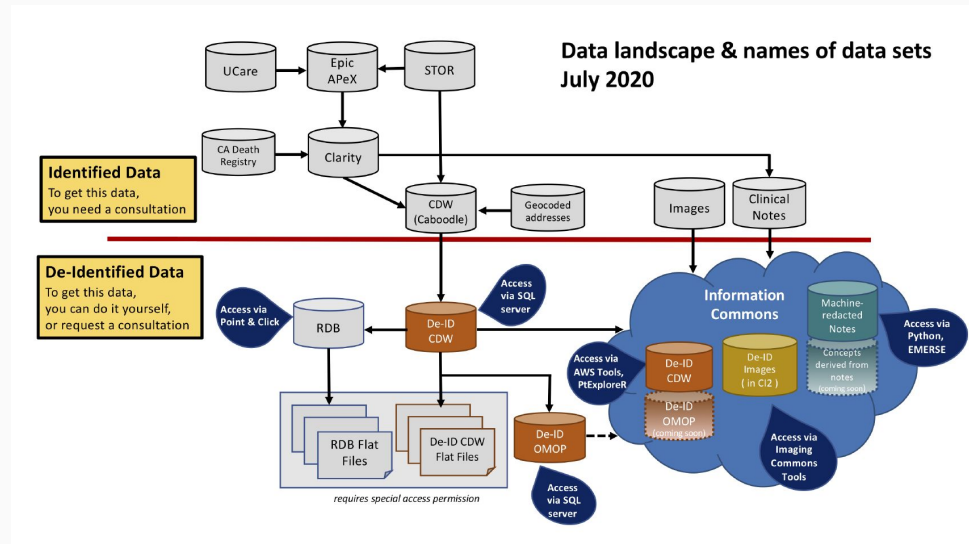
# Use Case

**Comparative Analysis of Social Connections/Isolation and  
Stress Documentation in Structured and Unstructured  
Machine De-Identified Data using PatientExploreR and  
EMERSE**



# Main use cases for EMERSE

1. Bring a list of patient IDs into EMERSE,
  - a. E.g. PatientExploreR, De-ID CDW queries, etc.
2. Search within EMERSE to create a list of patient IDs
3. Reviewing clinical notes using basic or advanced natural language processing (NLP) features



# EMERSE: Use-case example


- Comparative Analysis of Social Connections/Isolation and Stress Documentation in Structured and Unstructured Machine De-Identified Data using PatientExploreR and EMERSE
- Social determinants of health (SDOH) embedded in the clinical notes can be captured using NLP, but a corpi are hard to generate and data access can be challenging, often requiring advanced programming skills.

# EMERSE: Building Patient Cohorts

- We identified 4,274 adult ( $\geq 18$  years) patients receiving primary care for their diabetes (type I and type II: ICD-10-CM: E10, E11) at UCSF from January 1, 2018, to December 31, 2019.
  - Patient cohort was built using PatientExploreR

The screenshot shows the EMERSE web application interface. At the top, the header displays the EMERSE logo and the user name 'Shivani Mehta'. Below the header, there's a navigation bar with tabs: 'Patients' (showing 'All (2,508,721)'), 'Filters', 'Terms', and 'Results'. The 'Results' section has buttons for 'HIGHLIGHT DOCUMENTS' and 'FIND PATIENTS'. Below this, there's a section for 'Temporary Patient List' with tabs: 'Temporary Patient List', 'Saved Patient Lists' (highlighted with a red box), 'All Local Patients', and 'Network'. The 'Temporary Patient List' section has a sidebar with options: 'Name/Description', 'Add/Upload Patients' (highlighted), 'View Patients', 'Patient Demographics', 'Sharing', 'Clear', and 'Export'. The main content area shows instructions for adding patients and a form for 'Add Patients' with a text input for 'Medical Record Numbers' (highlighted with a red box) and an 'ADD PATIENTS' button. To the right, there's a section for 'Add by Bulk Upload' with a 'Choose File' button and an 'UPLOAD FILE' button.

# EMERSE: Search Filters



Patients

All (2,508,721)

Filters

Terms

Results

HIGHLIGHT DOCUMENTS

FIND PATIENTS

FILTERS

Encounter Dates

Age (current)

Sex

Race

Ethnicity

Vital Status


Document Source

Only documents dated between the following dates will be found in search and when reviewing patients' charts.

The date range for selection is from 02/01/2008 to 01/03/2022


Start Date:

MM/DD/YYYY



End Date:

MM/DD/YYYY



# EMERSE: Basic Search Terms

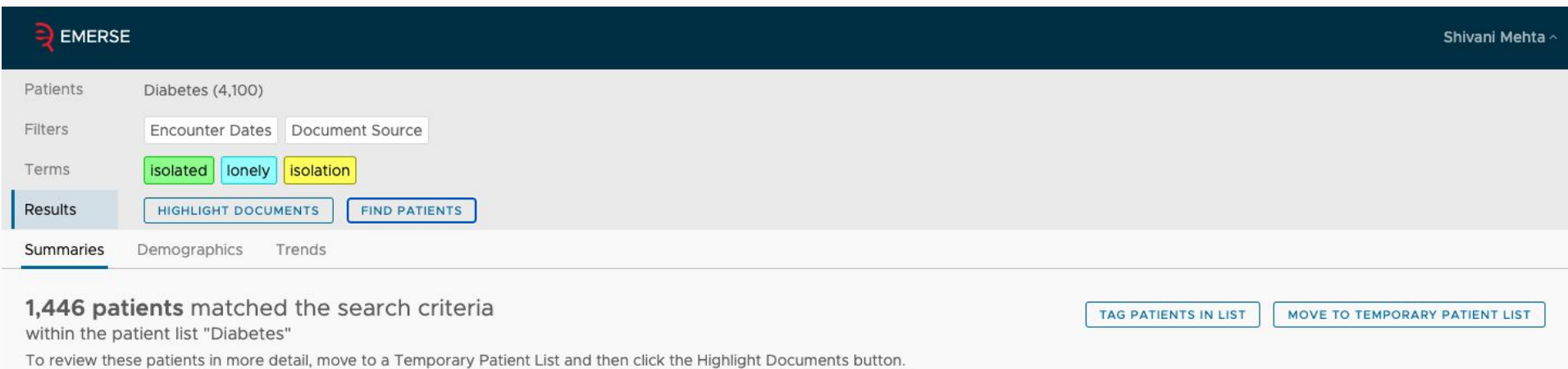
The screenshot displays the EMERSE web application interface. At the top, a dark blue header contains the EMERSE logo on the left and the user name "Shivani Mehta" on the right. Below the header, a light gray navigation bar shows "Patients All (2,508,721)" and "Filters". Under "Filters", the "Terms" tab is active, displaying three colored tags: "isolated" (green), "lonely" (cyan), and "isolation" (yellow). Below the tags are buttons for "HIGHLIGHT DOCUMENTS" and "FIND PATIENTS".

The main content area is divided into two columns. The left column contains a sidebar with the following links: "Temporary Terms", "Saved Terms", "Advanced Terms", "Temporary Terms", "Name/Description", "Manage Terms" (highlighted), "View Terms", "Sharing", and "Clear/Delete".

The right column contains two main sections. The top section, "Terms to include", has a sub-header "Please add terms or multi-word phrases, one at a time using the field below." It features a red-bordered input box with the label "Add term", a text input field, and an "ADD" button. Below this, a list of "Included terms" is shown: "isolated" (green), "lonely" (cyan), and "isolation" (yellow), each with a small edit icon and a dropdown arrow. The bottom section, "Phrases to exclude", has a sub-header "Please add phrases to exclude one at a time using the field below." It features a white-bordered input box with the label "Add Phrase", a text input field, and an "ADD" button. Below this, a section for "Excluded phrases" shows "None".

At the top right of the main content area, there are two buttons: "UPLOAD TERMS" and "SAVE TERMS".

# EMERSE: Basic Search Terms Results



The screenshot displays the EMERSE web application interface. At the top, the EMERSE logo is on the left, and the user name 'Shivani Mehta' with a dropdown arrow is on the right. Below the header, the main content area is divided into sections. The 'Patients' section shows 'Diabetes (4,100)'. The 'Filters' section includes 'Encounter Dates' and 'Document Source'. The 'Terms' section shows three colored buttons: 'isolated' (green), 'lonely' (cyan), and 'isolation' (yellow). The 'Results' section is highlighted with a blue bar and contains two buttons: 'HIGHLIGHT DOCUMENTS' and 'FIND PATIENTS'. Below the 'Results' section, there are tabs for 'Summaries', 'Demographics', and 'Trends'. The 'Summaries' tab is active, showing the text: '1,446 patients matched the search criteria within the patient list "Diabetes"'. To the right of this text are two buttons: 'TAG PATIENTS IN LIST' and 'MOVE TO TEMPORARY PATIENT LIST'. At the bottom, a note states: 'To review these patients in more detail, move to a Temporary Patient List and then click the Highlight Documents button.'

EMERSE Shivani Mehta ^

Patients Diabetes (4,100)

Filters Encounter Dates Document Source

Terms isolated lonely isolation

Results HIGHLIGHT DOCUMENTS FIND PATIENTS


Summaries Demographics Trends

**1,446 patients** matched the search criteria within the patient list "Diabetes"

Tag Patients in List Move to Temporary Patient List

To review these patients in more detail, move to a Temporary Patient List and then click the Highlight Documents button.

# EMERSE: Basic Search Terms Results

 EMERSE

Patients

Diabetes (4,100)

Filters

Encounter Dates

Document Source

Terms

isolated

lonely

isolation

Results

HIGHLIGHT DOCUMENTS

FIND PATIENTS

Summaries

Demographics

Trends

(DEID CDW)

01/29/2019

...procedure all the PVs were isolated. He had an SVC atrial...

...so we performed an SVC isolation. We also ablated a high...

...This patient had a PV isolation and atrial flutter ablation on...

...avg rate of 117 bpm. - Isolated \*\*\*\*\* were frequent (6.2%, \*\*\*\*\*), SVE...

...were rare(isolated VEs were rare (...)

...well s/p redo PV isolation in 2014. He has a...



# EMERSE: Basic Search Terms Results



The screenshot displays the EMERSE web application interface. At the top, the EMERSE logo is visible. Below it, the 'Patients' section shows 'Diabetes (4,100)'. The 'Filters' section includes 'Encounter Dates' and 'Document Source'. The 'Terms' section shows three selected terms: 'isolated' (green), 'lonely' (cyan), and 'isolation' (yellow). The 'Results' section is active, showing two buttons: 'HIGHLIGHT DOCUMENTS' and 'FIND PATIENTS'. Below the 'Results' section, there are tabs for 'Summaries', 'Demographics', and 'Trends'. The 'Summaries' tab is selected, showing a list of results. The first result is '(DEID CDW)' dated '02/03/2018'. The text of the result is: '...Transmission-based precautions: DROPLET Microorganism isolated: Rhinovirus/Enterovirus In addition to... standard precautions, please maintain DROPLET isolation for the duration of Mr...'. The words 'isolated' and 'isolation' are highlighted in green and yellow respectively.

EMERSE

Patients Diabetes (4,100)

Filters Encounter Dates Document Source

Terms isolated lonely isolation

Results HIGHLIGHT DOCUMENTS FIND PATIENTS

Summaries Demographics Trends

(DEID CDW) 02/03/2018

...Transmission-based precautions: DROPLET Microorganism isolated: Rhinovirus/Enterovirus In addition to...  
...standard precautions, please maintain DROPLET isolation for the duration of Mr...

# EMERSE: Advanced Search

- Advanced terms search feature offers precise control of searches
- Uses underlying JavaCC Lucene package, a search engine query parser
- Allows for many types of searches including nested Boolean operators, wildcard searches, proximity searches, fuzzy searches, term boosting
  - [https://lucene.apache.org/core/7\\_2\\_1/queryparser/org/apache/lucene/queryparser/classic/package-summary.html#Overview](https://lucene.apache.org/core/7_2_1/queryparser/org/apache/lucene/queryparser/classic/package-summary.html#Overview)

# EMERSE: Application of Advanced Search

SDOH DOMAIN	SUBDOMAIN (ICD-10 codes)	TERMS
Social connections / isolation	<i>Social Exclusion and Rejection</i> (Z60.4, R45.8, R45.89, Z65.9, Z91.8, Z91.89, Z65.9, Z91.8)	("social isolation"~10) OR ("socially isolated"~10) OR ("Lonely", "Loneliness") NOT ("no loneliness", "not lonely", "no social isolation")
	<i>Social isolation prompted by family-related disruptions</i> (Z60.2, Z63.0, Z63.5, Z63.9)	
	<i>Lack of social environment and social support</i> (Z60.8, Z63.8, Z73.4)	
Stress	<i>General terms</i> (F43.2, F43.20, F43.9, Z73.3, Z73.2)	("stress", stressor*, "stressful") OR ("feeling stressed"~10) NOT (stress scor*, "stress lab", "echocardiogram", "ecg", stress imag*, stress test*, "stress myocardial perfusion", "stress ulcer", "post traumatic stress disorder", "stress induced", "stress reduction", "stress fracture", "stress reaction", "pharmacologic stress", "post stress", "stress dose", "exercise stress", "no stress", "stress induce ischemia", "stress echo") NOT ("stress study"~5) NOT ("stress calculated"~5)
	<i>Work-related stress</i> (Z56.3, Z56.6, Z63.7)	
	<i>Family/home-related stress</i> (Z63.7, Z63.79, Z63.8)	
	<i>Life-management related stress</i> (Z73.8, Z73.89, Z91.8, Z91.89)	

# EMERSE: Application of Advanced Search

The screenshot displays the EMERSE application interface. At the top, the EMERSE logo is visible. Below it, the 'Patients' section shows 'Diabetes (4,100)'. The 'Filters' section includes 'Encounter Dates' and 'Document Source'. The 'Terms' section contains a complex search query: `("social isolation"-10) OR ("socially isolated"-10) OR ("Lonely", "Loneliness") NOT ("no loneliness", "not lonely", "no social isolation")`. The 'Results' section has two buttons: 'HIGHLIGHT DOCUMENTS' and 'FIND PATIENTS'. Below the results section, there are three tabs: 'Summaries', 'Demographics', and 'Trends'. The 'Summaries' tab is active, showing two search results. Each result includes a patient identifier '(DEID CDW)', a date, and a snippet of text with highlighted terms.

EMERSE

Patients Diabetes (4,100)

Filters Encounter Dates Document Source

Terms `("social isolation"-10) OR ("socially isolated"-10) OR ("Lonely", "Loneliness") NOT ("no loneliness", "not lonely", "no social isolation")`

Results HIGHLIGHT DOCUMENTS FIND PATIENTS

Summaries Demographics Trends

(DEID CDW) 07/10/2018

...Pt is **lonely**, going to move back to \*\*\*\*\* where she will have more family and can afford...

(DEID CDW) 02/14/2018

... around wanting more to do and feeling **lonely**. Discussed behavioral activation as method of addressing...  
... mild/moderate anxiety, especially in setting of **loneliness** and \*\*\*\*\* up with plan to patient to...  
... go to library daily in morning to help address feeling of isolation and **loneliness**, especially given...  
..., anxiety, depression, feelings of **loneliness**, and psychosocial stressors including relationship...


# EMERSE: Summary of Search Options

**Table 3. Feature Comparison between the *Temporary Terms*, the *Saved Terms*, and the *Advanced Terms***

Feature	Temporary Terms	Saved Terms	Advanced Terms
Maximum number of terms per list	no limit <sup>1</sup>	no limit <sup>1</sup>	no limit <sup>1</sup>
Supports 18 colors for terms	Yes	Yes	No
Full control over the choice of colors for terms	Yes	Yes	No
Ideal for teams working together	No	Yes	No
Supports sharing with other users	No	Yes	No
Saved between sessions	No	Yes	No
Has a Term Upload feature	Yes	Yes	No
Can take advantage of Synonym suggestions	Yes	Yes	No
Spell Checking of Terms	Yes	Yes	No
Supports the ability to search for terms in a case-sensitive manner	Yes	Yes	Yes
Supports advanced search features including customizable proximity searches, fuzzy searches, Boolean searches (AND, OR NOT), and regular expressions	No	No	Yes

<sup>1</sup> While there is no limit on the number of terms, a large number (more than 100) will cause the system to run quite slow.

# EMERSE: Highlighting Documents and Tagging Patients

 EMERSE

Shivani Mehta ~

Patients Diabetes (4,100)

Filters Encounter Dates Document Source
















Terms ("social isolation"-10) OR ("socially isolated"-10) OR ("Lonely", "Loneliness") NOT ("no loneliness", "not lonely", "no social isolation")

Results HIGHLIGHT DOCUMENTS FIND PATIENTS

Overview

Sorted by: Insert Order

Numbers Grayscale Mosaic

MRN	Name	DEID CDW	Main EHR	Pathology	Other	Radiology	Comment	Tag
D383440D2FFBD2	DEID, DEID DEID						 0 / 255	
D445248DBC9ICA	DEID, DEID DEID						 0 / 255	
DF4290441F77E1	DEID, DEID DEID						 0 / 255	
D69FD215E09CE5	DEID, DEID DEID						 0 / 255	
DB11B6756DC571	DEID, DEID DEID						 0 / 255	
DB5F714050A6DD	DEID, DEID DEID						 0 / 255	

Patients per page 50 1 - 50 of 4100 patients |< < 1 / 82 > >|

# EMERSE: Exporting Patient Cohorts

The screenshot displays the EMERSE web application interface. At the top, the EMERSE logo is visible. The main section is titled "Patients" and shows a cohort of "Diabetes (4,100)". Below this, there are filter options for "Encounter Dates" and "Document Source". The "Terms" field contains a complex query: `("social isolation"-10) OR ("socially isolated"-10) OR ("Lonely", "Loneliness") NOT ("no loneliness", "not lonely"-10, "no social isolation")`. The "Results" section has two buttons: "HIGHLIGHT DOCUMENTS" and "FIND PATIENTS".

Below the main section, there are tabs for "Temporary Patient List", "Saved Patient Lists" (which is selected), "All Local Patients", and "Network". Under the "Saved Patient Lists" tab, there is a list of saved lists, including "Diabetes".

On the left side, there is a sidebar menu with options: "Name/Description", "Add/Upload Patients", "View Patients", "Patient Demographics", "Sharing", "Clear/Delete", and "Export" (which is highlighted).

The main content area shows a form for adding a password to protect the Excel file. It includes fields for "Enter password" and "Re-enter password", both with a strength indicator (a green bar) and a "CLEAR" button. A message states: "A strong password is required".



# Additional Resources

- For access
  - <https://wiki.library.ucsf.edu/pages/viewpage.action?spaceKey=UCNFR&title=EMERSE>
- Navigating EMERSE
  - <https://ucsf.app.box.com/s/x3ulqxupj6dlk5ttf3pixsggzckv6zd>
- Training resources from UMich
  - <http://project-emerse.org/tutorials.html>
  - [http://project-emerse.org/documentation/user\\_guide.html](http://project-emerse.org/documentation/user_guide.html)

# Questions?



# How to get permission

- Active **CAMPUS** or **SOM** account.
- If user has a UCSFMC account, a new CAMPUS account can be requested via Account Request Form [https://ucsf.service-now.com/ucsfit?id=ucsf\\_sc\\_cat\\_item&sys\\_id=dcd937132b779a0019d7c71317da151f](https://ucsf.service-now.com/ucsfit?id=ucsf_sc_cat_item&sys_id=dcd937132b779a0019d7c71317da151f)

Request Instruction can be found on the wiki <https://wiki.library.ucsf.edu/x/UitXHw>

# Questions?



# Audience guided interactive session



# Questions?



# User Friendly Natural Language Processing (NLP) of Free-text Clinical Notes at UCSF: EMERSE and cTAKES



William Brown III, PhD, DrPH, MA

Shivani Mehta, MPH



# IRB Approval for Machine-Redacted Notes Research

Machine Redacted Notes are clinical notes extracted from UCSF EPIC, which have been processed by software to remove PHI. The data extracted includes clinical text for any patient who has UCSF EMR records. Clinical notes are extracted from EPIC, and then undergo machine redaction (PHI de-identification software), for removal of 18 PHI elements as identified by HIPAA. Dates are not removed, but instead shifted for each patient following the same methodology as de-identified Clinical Data Warehouse (de-id CDW). Each patient's records are assigned synthetic identifiers, following de-id CDW process, which enables linking notes to structured data in de-id CDW.

- The process of machine redaction and resulting data are currently undergoing formal certification by third party vendor. While this process is time consuming, "Machine-redacted pre-certified data for Machine Learning and Analysis" UCSF IRB Study #19- 27988 was in put in place with the aims to (1) speed-up access time to machine-redacted clinical data for research purposes and (2) to gain researcher feedback. Until the database receives de-identification certification, these data are treated as identified data. The data are stored in designated secure environments in IT data center. Researchers can access the data via secure software end-points, that require user authentication. The software enables researchers to query and analyze the data, download data and combine with other data sets.

- If researchers include the use of Machine Redacted Notes for purposes of their IRB-approved study, it is researcher responsibility to (1) limit data analyses to those approved for their study, (2) use the minimum data necessary for their study, (3) combine machine redacted data sets with other data only as approved by IRB for their study, (3) securely transfer and store downloaded data as approved by their IRB study, (4) not attempt to re-identify patients (5) comply with data sharing practices as approved by their IRB study and UCSF Data Sharing task force.



# Access Requirements

- Copy of approved IRB document
- Users requesting access should be listed as study personnel
- IRB text allowing Machine-Redacted Notes research <https://wiki.library.ucsf.edu/x/BwRXHw>
- **Active CAMPUS or SOM account.**
- If user has a UCSFMC account, a new CAMPUS account can be requested via Account Request Form  
[https://ucsf.service-now.com/ucsfit?id=ucsf\\_sc\\_cat\\_item&sys\\_id=dcd937132b779a0019d7c71317da151f](https://ucsf.service-now.com/ucsfit?id=ucsf_sc_cat_item&sys_id=dcd937132b779a0019d7c71317da151f)

Request Instruction can be found on the wiki

# EMERSE

## Machine-Redacted Notes

Clinical notes extracted from UCSF Electronic Health Record, Epic, and processed by software to remove 18 PHI elements. Each patient's records are assigned synthetic identifiers that maintain the link between notes and the patient's structured data in DEID CDW. There are two datasets of UCSF Machine Redacted Notes, known as 1) General Set and 2) COVID-19 Set.

## EMERSE: **E**lectronic **M**edical **R**ecord **S**earch **E**ngine

An open-source software tool to search and browse notes data, without programming experience. UCSF has installed two instances of EMERSE here at UCSF against our 1) machine-redacted notes and 2) COVID-19 notes. EMERSE was created by the University of Michigan.