UCSF Information Commons and Cancer Commons

Sharat Israni, PhD, CTO & Exec Director & Gundolf Schenk, PhD, Principal Data Scientist

NCI Containers & Workflows Interest Group
June 10, 2022

A joint effort of BCHSI/CTSI/Radiology/UCSF-IT/UCSF-Library
Deck substantially authored by Albert Lee, MS, BCHSI & UCSF-Library

UCSF Bakar Computational Health Sciences Institute

What is the Opportunity Here?

Today's integrative research and precision medicine, need the **new generation of insights** …

… afforded by **wide and deep multifactor data**,

made possible by **data science, AI and intelligent computing power**, and

demanded by doctors, researchers and patients

Information Commons

Everything you need for translational data science in one place

Data

Tools

Compute

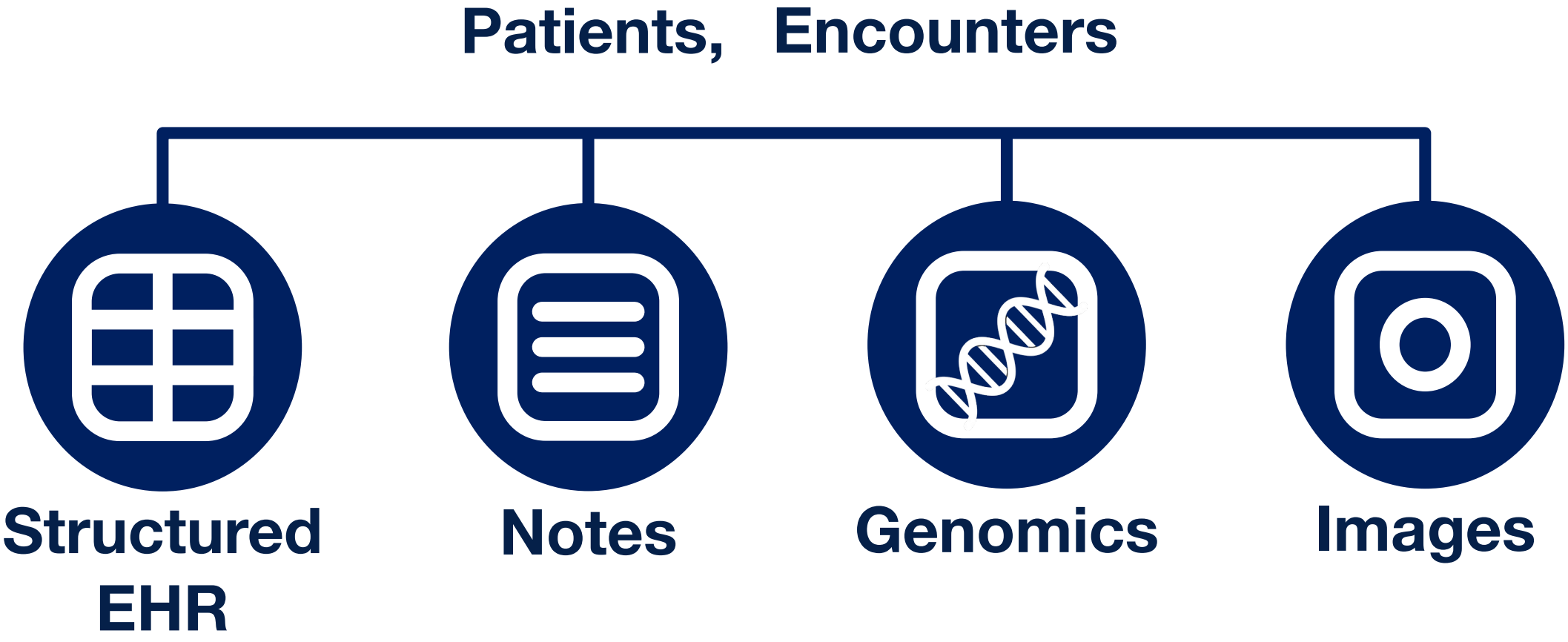
Community

Without a further IRB

Data

Data

Shared multi-modal, de-identified, linked clinical data



Structured EHR Data

Large dataset accessible through multiple schemas.

Available in both CDW and OMOP formats

- 5.5M Patients
- 100M Encounters
- 200M Diagnoses
- 375M Lab Results
- …

Structured EHR Data

Access compatible across the University of California System

UC DAVIS MEDICAL CENTER

UCLA Health

UC San Diego Health System

UC Irvine Health

UC Riverside

UCSF Health

Clinical Notes and Structured Extracts

Searchable text de-identified through machine redaction.

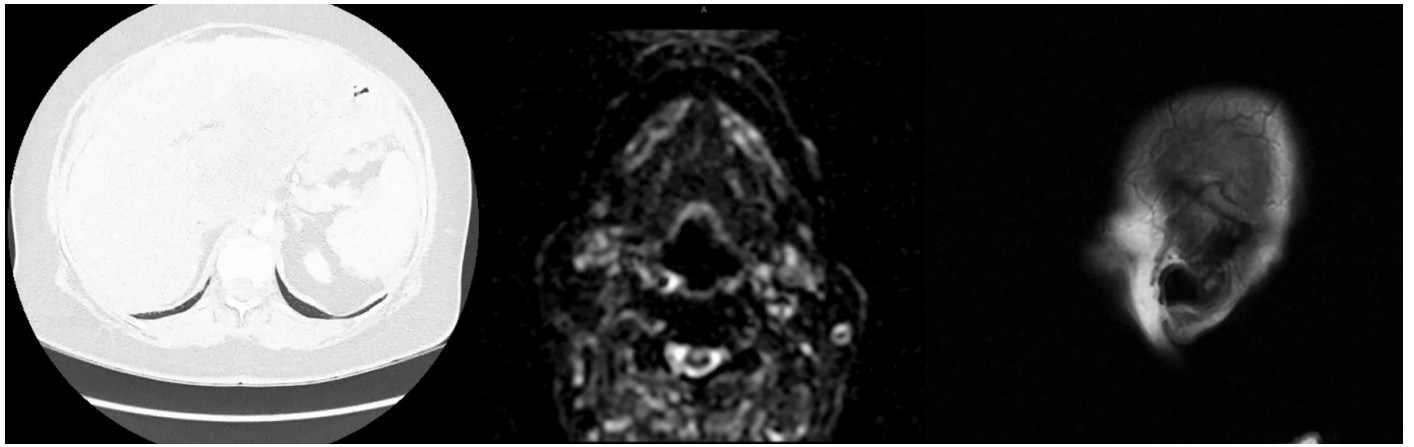
Included with access to DeID Research Data Assets

- Over 115M Notes
- 4.9B extracted clinical concepts

Imaging Commons

Searchable images with de-identified metadata.

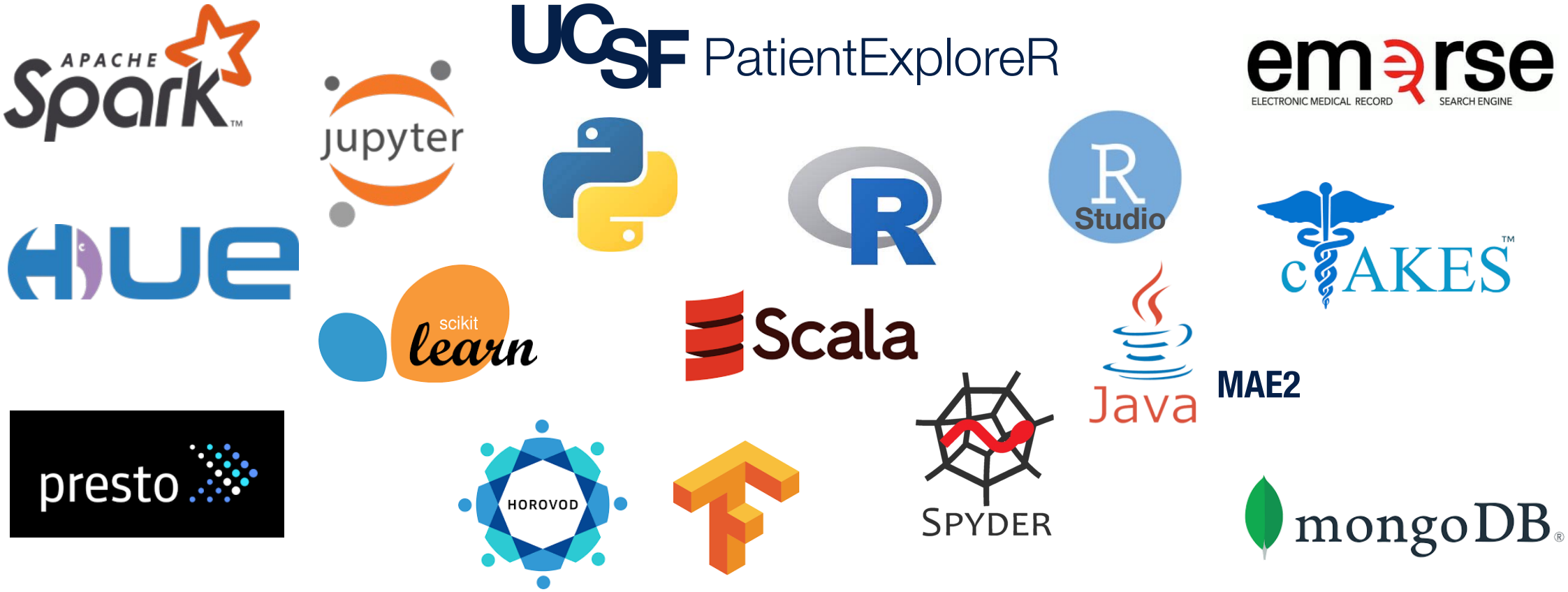
Over 7.5M Images!



Tools

Tools

Applications to support every part of data research



Data Exploration

User-friendly interfaces for quick data exploration

Explore

Structured EHR

Search

Clinical Notes

Browse

Medical Images

Code Development

Popular IDEs to support code development

Python

R

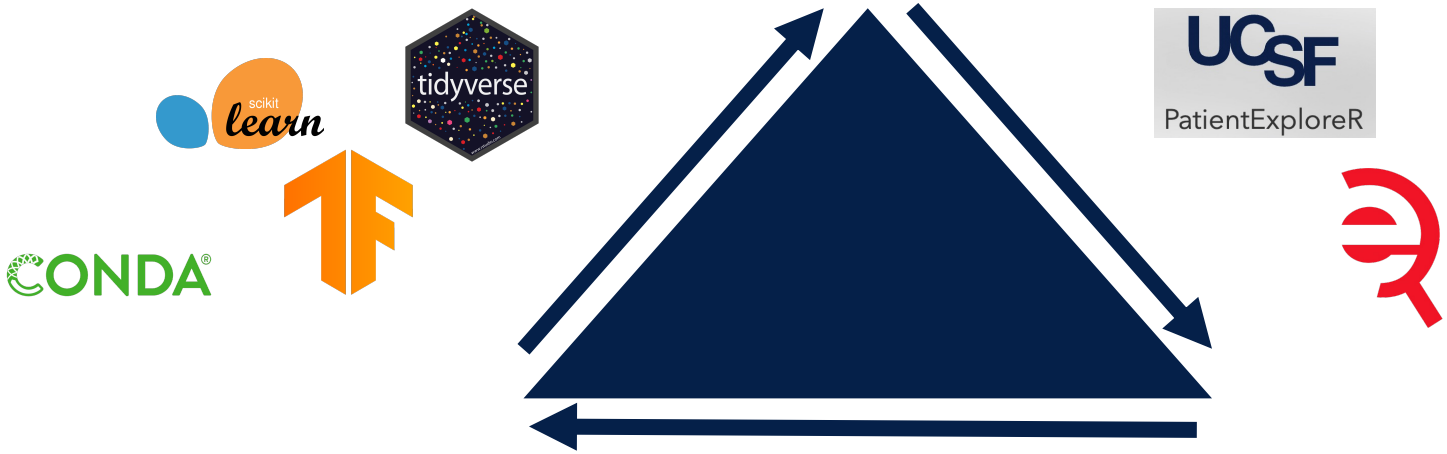
SQL

Prevalent Libraries for Data Science and Machine Learning

Tools

Applications to support every part of data research.

Data Exploration



Data Analysis

Data Extraction

Compute

Compute

Computing environments enable complex data analysis.

On a Large Premise Cluster

On AWS Cloud

As a Collection of Customizable Virtual Machines

Community

Community

Connecting people and knowledge

Documentation

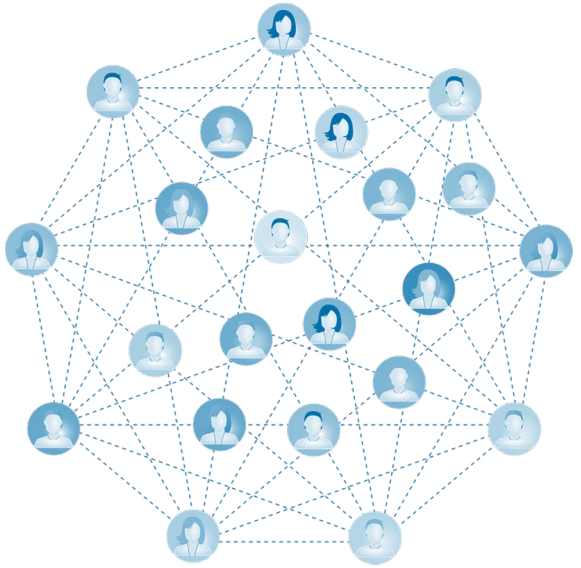
Tutorials and guides

User Support

Weekly office hours and Slack channels

Meetups

Monthly meetups (e.g. NLP@UCSF)



I would not be able to propose any of these projects without [this deidentified] infrastructure and tools. My K23 and AHA career development award are heavily supported by knowledge gained from these databases….. All my collaborators from other top universities in the country, which boast having great access to data and resources, are jealous of the setup we have here.

MD, Associate Prof. Clin. Neurology, UCSF and ZSFG

“

Information commons and its promise to make hypothesis testing on real world clinical data easier …was the reason I came across the country to do my clinical training at UCSF. … I have only just begun the research phase of my training, but the resource has exceeded even my optimistic expectations.

In no small part this is due to the dedicated team at the Bakar Institute that constantly is working to help researchers ask their question and making sure all resources are constantly being improved and as accessible as possible.

MD, PhD, Clinical Fellow, Medicine

“

We are in …planning a clinical pipeline for knee injury using multi-modal data (EHR + Imaging). As a team we have benefitted from practically all the resources available through IC, including databases, computational environments, and no less important, support.

Data Scientist, Dept. of Radiology & Imaging Sciences

Predicting Hospital Acquired Sepsis Onset Using the Information Commons

Xinran (Leo) Liu
Assistant Clinical Professor, Division of Hospital Medicine, UCSF
Director of Clinical Informatics at SMMC
Associate Program Director, Clinical Informatics Fellowship

UCSF

Harnessing the Power of EHR Data to Advance the Study of Interstitial Lung Disease

Erica Farrand MD

Assistant Professor, Division of Pulmonary, Critical Care, Sleep, Allergy and Sleep Medicine

UCSF

EXTRACTING INFORMATION from CLINICAL NOTES

A. Molinaro, PhD

Getting critical marker values and doctor sentiment for gliomas and glioblastomas

A. Odisho, MD, MPH, B. Yu,* PhD

Extracting grades, metastases and more for urothelial cancers

* UC Berkeley

UCSF

Survival Analysis of Primary CNS Lymphoma Using De-ID Clinical Data and Images

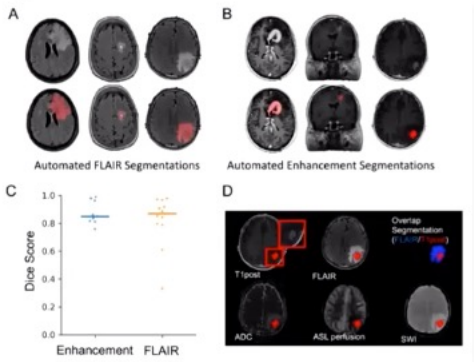
Reza Eghbali, PhD
Innovate for Health Fellow
UCSF and UC Berkeley
reza.eghbali@ucsf.edu
eghbali@berkeley.edu

Andreas Rauschecker MD, PhD
Assistant Professor
Radiology & Biomedical Imaging
UCSF

Aims: Help the radiologist and the oncologist

Automated Radiographic Treatment Response

Evaluate the accuracy of an **automated radiographic response to treatment evaluation** in PCNSL



Patient Stratification (Prognosis)

Identify low and high risk patients using quantitative **image characteristics** combined with key **clinical characteristics**

- ❏ Apparent Diffusion Coefficient imaging has been shown to be informative for treatment response [Barajas et al., '10; Huang et al, '10; Baek et al. '20]
- ❏ Hypothesis: Combining imaging and clinical biomarkers improves accuracy of prognosis.

University of California
San Francisco

informationcommons.ucsf.edu

BCHSI
CTSI
Radiology
UCSF IT
UCSF Library