

Automating the Compliance of Health Research with Virtual Research Desktop (VRD)

Nguyen Trieu, Associate Director, Biomedical Informatics, UC San Diego Health

Mike Hogarth, MD, Clinical Research Information Officer, UC San Diego Health


Andrew Greaves, Enterprise Cloud Architect, UC San Diego Health

Jit Bhattacharya, CEO/Founder, Xpertech Solutions

Why Is This Topic Important?


- Inappropriate removal of UC research data assets.
- Inappropriate third party data transfer requests ‘and the IRB approved it...’
 - Human research protections (IRB) vs policy and legal compliance (Data Use Agreement)
- A common risk for data breaches are loss of laptop/mobile devices.
 - Sutter hospital loss
- Ransomware
 - Unmanaged, unprotected, or misconfigured devices

USC to pay \$50 million and apologize to UC San Diego for poaching its Alzheimer's research program

Other Breaches  Sutter Health
With You. For Life.

In October 2011, Sutter Health reported the theft from its Sutter Medical Foundation of an **unencrypted** desktop computer containing information 4.2 million patients (see: **Computer Theft Affects 4.2 Million**). That incident resulted in the filing of 11 class action lawsuits. Those suits were **consolidated** into one case, which is making its way through Sacramento County Superior Court.

In addition, Sutter Health reported a May 2011 breach at its Sutter Gould Medical Foundation in which lost paper records resulted in 1,920 patients being notified that their information was possibly compromised. That incident appears on the Department of Health and Human Services **breach website** that lists incidents involving 500 or more individuals.

2017 Equifax data breach 

The Equifax data breach occurred between May and July 2017 at the American credit bureau Equifax. Private records of 147.9 million Americans, along with 15.2 million British citizens and about 19,000 Canadian citizens were compromised in the breach, making it one of the largest cybercrimes related to identity theft. [Wikipedia](#) exposed," said a UCSF statement [news release](#) on June 26. "The data that was encrypted is important to some of the academic work we pursue as a university serving the public good," continued the statement. "We therefore made the difficult decision to pay some portion of the ransom, approximately \$1.14 million, to the individuals behind the malware attack in exchange for a tool to unlock the encrypted data and the return of data they obtained."


The move also USC took conti and gave jobs t

CRYPTED
ther important
with unique key,
computer.

History and Compliance

Andrew Greaves, Enterprise Cloud Architect, UC San Diego Health
Mike Hogarth, MD, Clinical Research Information Officer, UC San Diego Health

What's The History?

- In 2017 we identified about 800 UCSD Health associated AWS accounts
 - We had ZERO visibility or security controls in place to monitor account activity.
 - We did not know what types of data, work loads, or potential risks associated.
 - UC San Diego Health did not have an AWS environment for research.
- Dr. Hogarth's first week at UCSD
 - Data Extraction of 10 million narrative clinical notes.
- “Build it (right) and they will come”
 - In late 2017 we started working with AWS Professional Services and Xpertech to help build out the [UCSD Health Secure Research Cloud \(HSRC\)](#).
 - We partnered with IS Security, ACTRI, DBMI, and UCSD Health research groups.

Quick Review of Federal and California Privacy Laws

- The Health Insurance Portability and Accountability Act (HIPAA), 1996
 - Electronic Protected Health Information (ePHI) and the 18 identifiable elements
 - HIPAA is a **policy**, not specific security controls. Only two specific technical controls are mentioned no generic logins and encryption required
 - **Covered entities** are defined in the **HIPAA** rules as (1) health plans, (2) health care clearinghouses, and (3) health care providers who electronically transmit any health information.
- California's Confidentiality of Medical Information Act (CMIA) 2009
 - Provides stronger privacy protections for medical information.
 - CMIA's primary purpose is to protect an individual's medical information, in electronic or paper format, from unauthorized disclosure.
 - **Personal** and Administrative Fines and Civil Penalties (including jail time)



Four Technical Safeguards Categories for PHI

1. Access Control

- A covered entity must implement technical policies and procedures that **allow only authorized persons** to access electronic protected health information (e-PHI)

2. Audit Controls

- A covered entity must implement hardware, software, and/or procedural mechanisms to record and examine access and other **activity in information systems** that contain or use e-PH

3. Integrity Controls

- A covered entity must implement policies and procedures to ensure that e-PHI is **not improperly altered or destroyed**. Electronic measures must be put in place to confirm that e-PHI has not been improperly altered or destroyed

4. Transmission Security

- A covered entity must implement technical security measures that **guard against unauthorized access** to e-PHI that is being transmitted over an electronic network.

Framework(s) for Achieving Compliance

- National Institute of Standards and Technology (**NIST**)
 - Maps to security controls detailed in NIST SP 800-53 (FISMA moderate)
 - Both technical and organizational (access controls) security controls
 - <https://www.nist.gov/healthcare/security/hipaa-security-rule>
- Center for Internet Security (**CIS**)
 - CIS Level1 meets FISMA moderate and HIPAA requirements
 - CIS resources are developed to work well as stand-alone resources or as companions to additional frameworks
 - <https://www.cisecurity.org/cybersecurity-tools/mapping-compliance/>

Moderate Impact Controls

Showing 128 controls

No.	Control	Priority	Low	Moderate	High
AC-1	ACCESS CONTROL POLICY AND PROCEDURES	P1	AC-1	AC-1	AC-1
AC-2	ACCOUNT MANAGEMENT	P1	AC-2	AC-2 (1) (2) (3) (4)	AC-2 (1) (2) (3) (4) (5) (11) (12) (13)
AC-3	ACCESS ENFORCEMENT	P1	AC-3	AC-3	AC-3
AC-4	INFORMATION FLOW ENFORCEMENT	P1	AC-4	AC-4	AC-4
AC-5	SEPARATION OF DUTIES	P1	AC-5	AC-5	AC-5
AC-6	LEAST PRIVILEGE	P1	AC-6 (1) (2) (5) (9) (10)	AC-6 (1) (2) (3) (5) (9) (10)	AC-6 (1) (2) (3) (5) (9) (10)
AC-7	UNSUCCESSFUL LOGIN ATTEMPTS	P2	AC-7	AC-7	AC-7
AC-8	SYSTEM USE NOTIFICATION	P1	AC-8	AC-8	AC-8
AC-11	SESSION LOCK	P3	AC-11 (1)	AC-11 (1)	AC-11 (1)
AC-12	SESSION TERMINATION	P3	AC-12	AC-12	AC-12

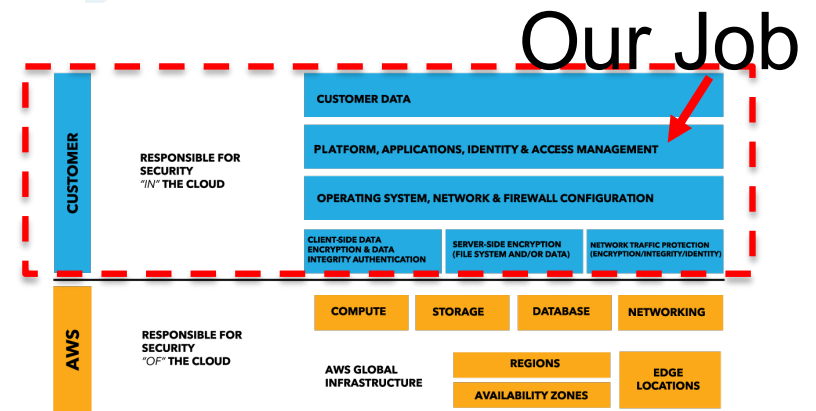


Compliance in AWS (PHI/HIPAA)

- Business Associate Agreement (BAA). Extension of the **covered entity** to vendor and contractors.
- Shared Responsibility Model
 - Hypervisor level and below is AWS responsibility
 - Above is customer responsibility
- HIPAA **eligible** ≠ compliant
 - Cloud providers offer ‘HIPAA eligible services’
 - This still requires customer to apply all controls that fall in their section of the shared responsibility model
 - Non-HIPAA compliant services can be used in the architecture as long as no PHI/PII data passes through those services

HIPAA Business Associate Agreements

The HIPAA Regulations reflect the understanding that a covered entity, such as the University of California, often requires the services of third parties (“business associates”) to conduct its operations. A business associate is a person or entity that creates, receives, maintains or transmits protected health information (“PHI”) on behalf of the University. A business associate relationship exists when an individual or entity, acting on behalf of the University, assists in the performance of a function, activity or service involving the use or disclosure of PHI. These functions, activities and services, to or on behalf of the covered entity, include, but



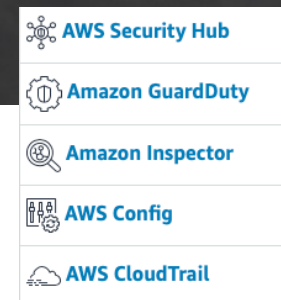
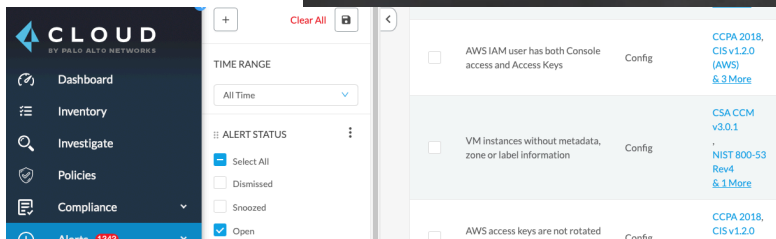
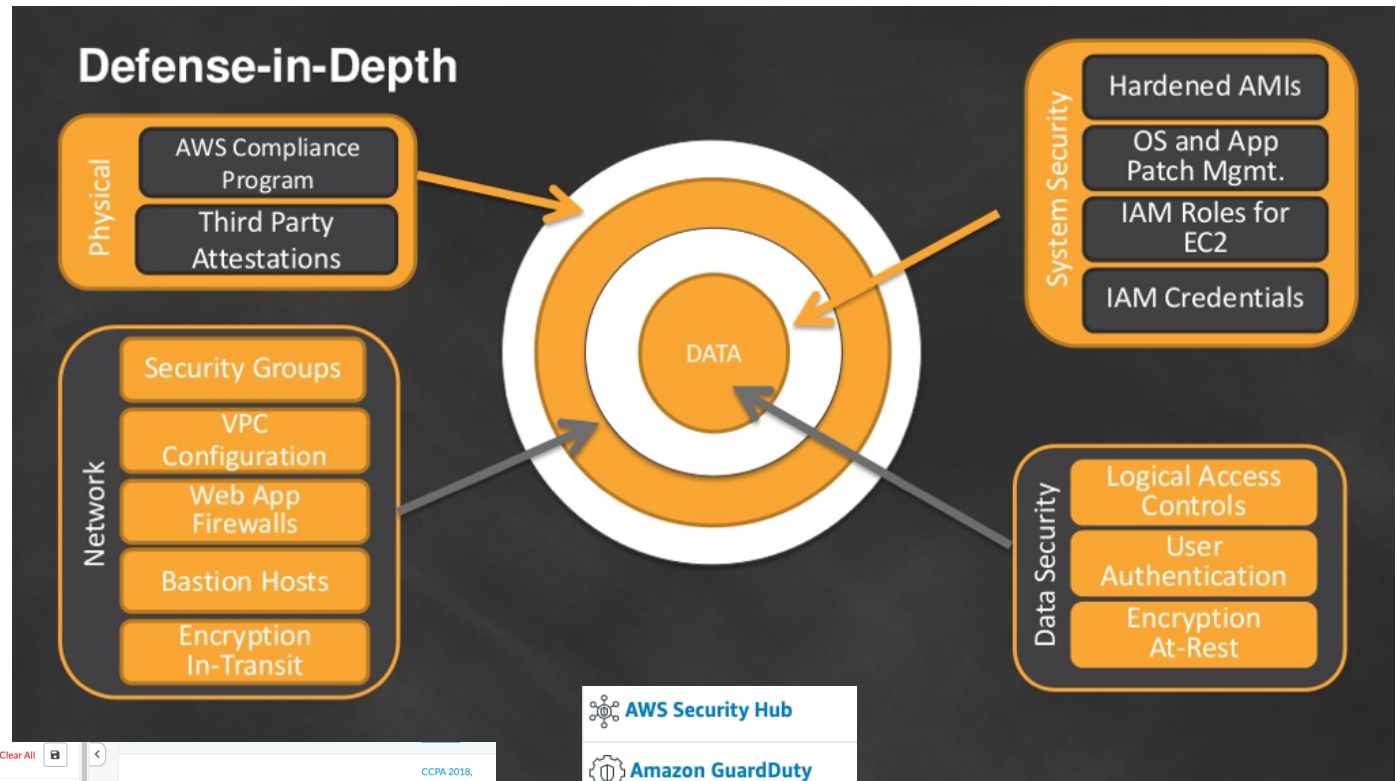
Architecture:

UCSD Health Secure Research Cloud (HSRC)

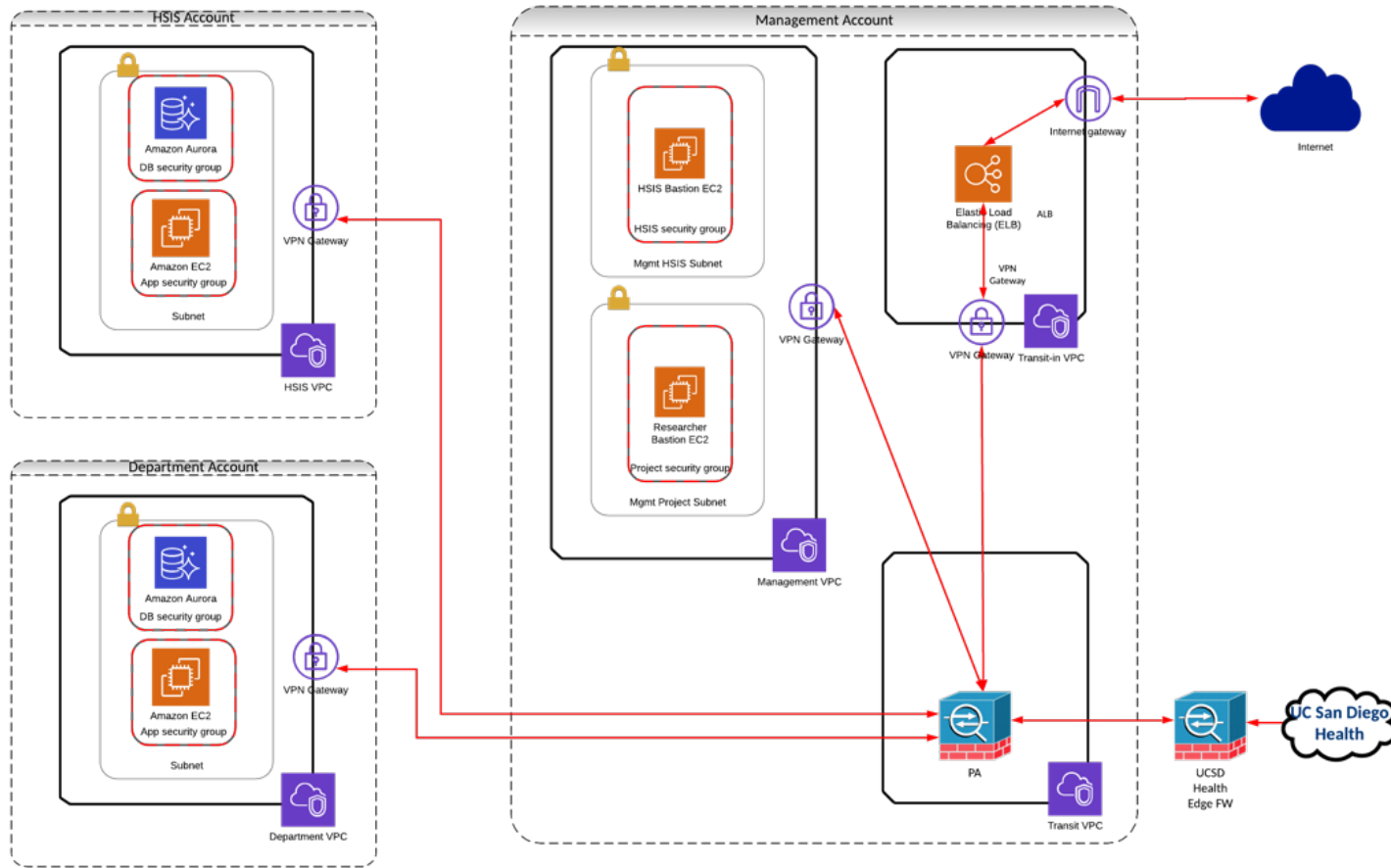
Jit Bhattacharya, CEO/Founder, Xpertech Solutions
Andrew Greaves, Enterprise Cloud Architect, UC San Diego Health

AWS HIPAA Environment

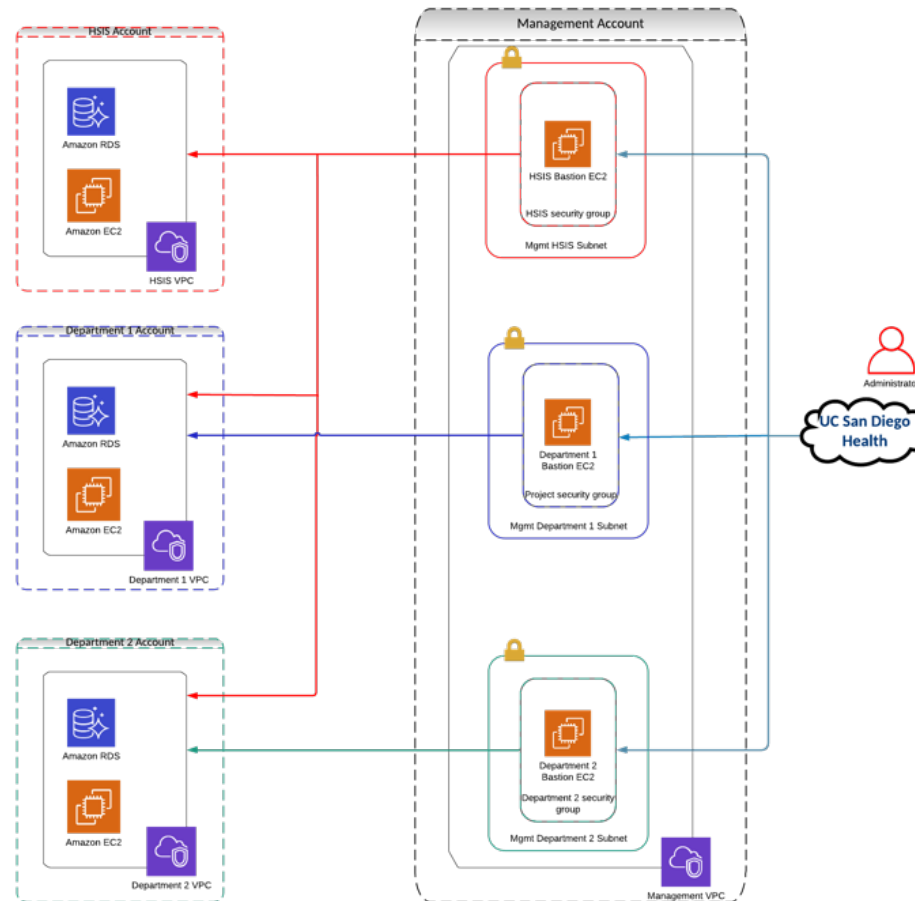
- Multi-layered Security Approach
- HIPAA Boundary
- Centralized Logging
 - VPC Flow Logs, CloudTrail, System logs
- Transit VPC and Palo Alto
- Ingress and Egress is only through the UCSDH PA's and their defined network rules
- VPC isolation
- CISO approval required



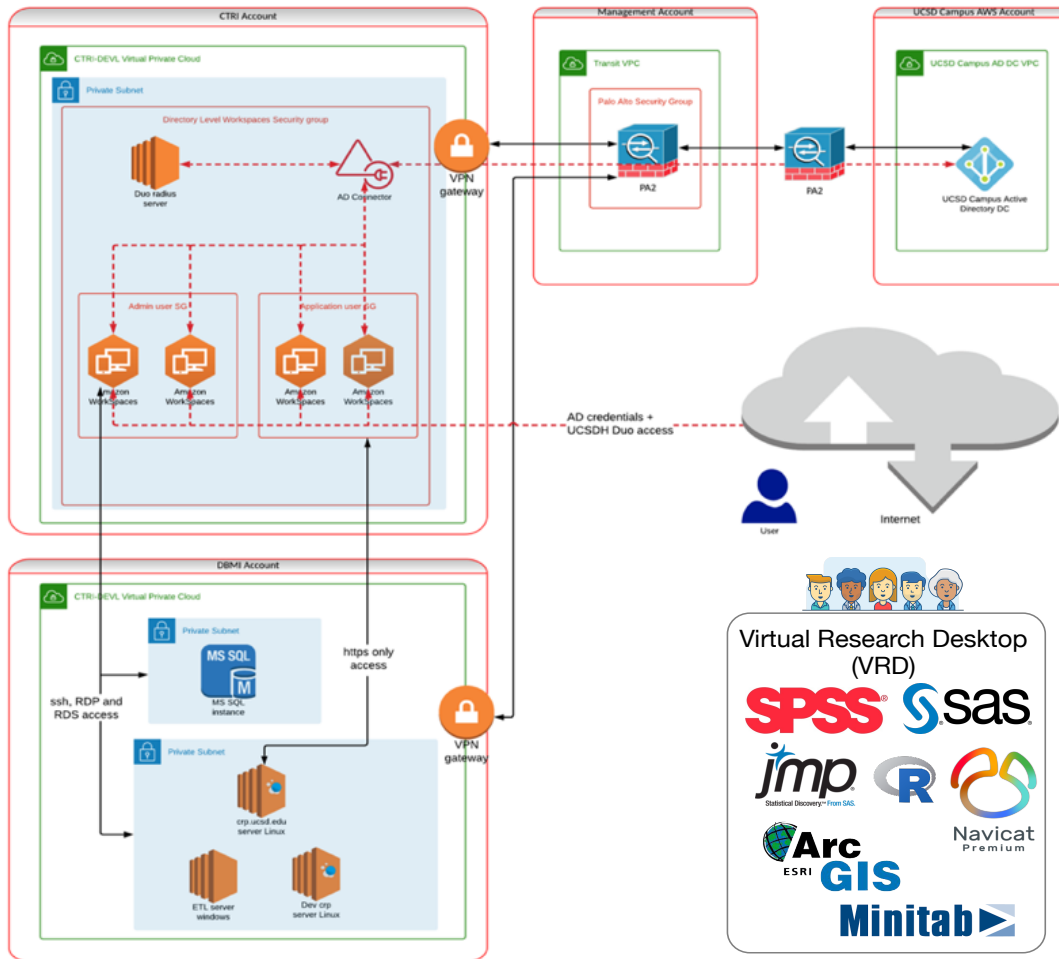
HIPAA Boundary Controls (simplified)



Multi-Researcher/Department Access Architecture



Secure AWS workspaces: Virtual Research Desktop (VRD)



Virtual Research Desktop (VRD)

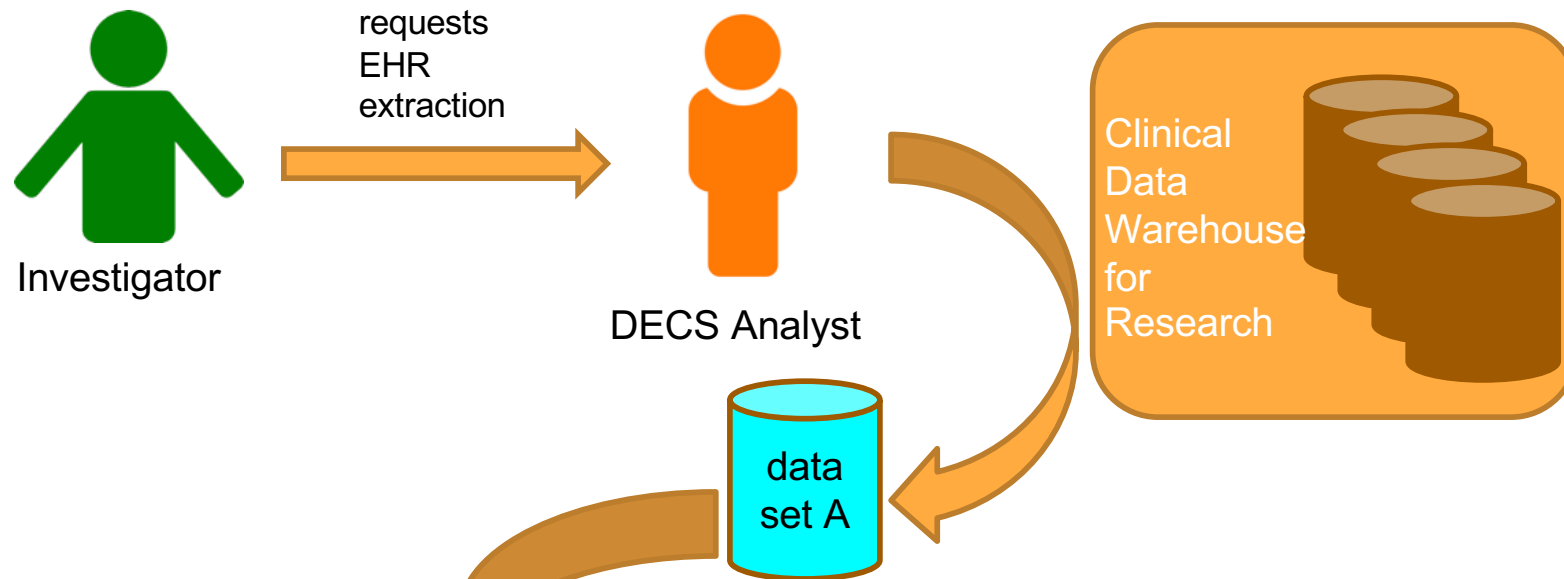
The VRD includes logos for the following software: SPSS, sas, jmp, R, ArcGIS, Minitab, and Navicat Premium.



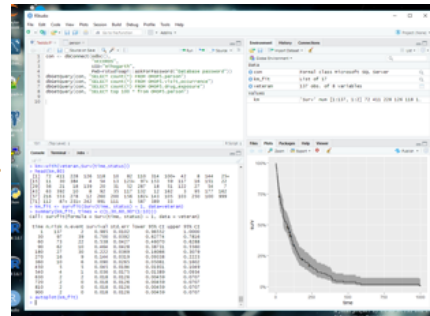
ACTRI, the VRD and Use Cases

Nguyen Trieu, Associate Director, Biomedical Informatics, UC San Diego Health
Mike Hogarth, MD, Clinical Research Information Officer, UC San Diego Health

ACTRI –Data Extraction Concierge Service (DECS) and VRDs



placed by Analyst into Investigator's VRD – "SecureDrop" directory



Virtual Research Desktop (VRD)

Configurations:

- Standard: Win10, 2 CPU, 4Gb mem
- Super User: Win 10, 8 CPU, 32Gb mem
- Software: Rstudio/R, Python/PyCharm, SPSS, tag.bio, DataGrip SQL tool, MATLAB, Java 8 JDK, MySQL WB, PgAdmin

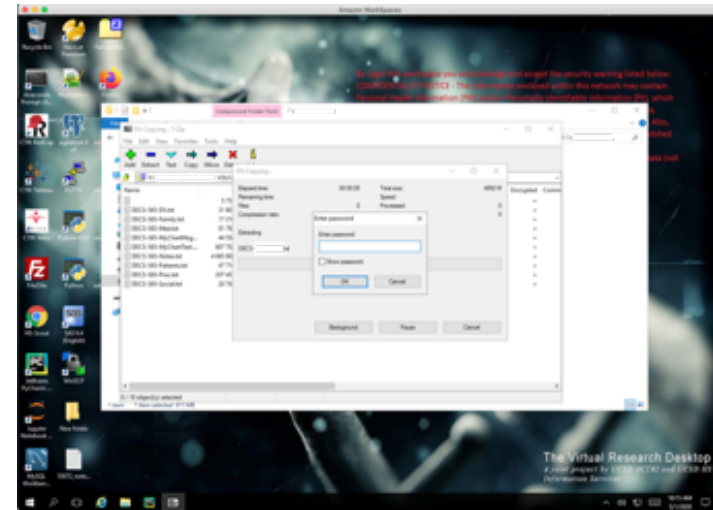
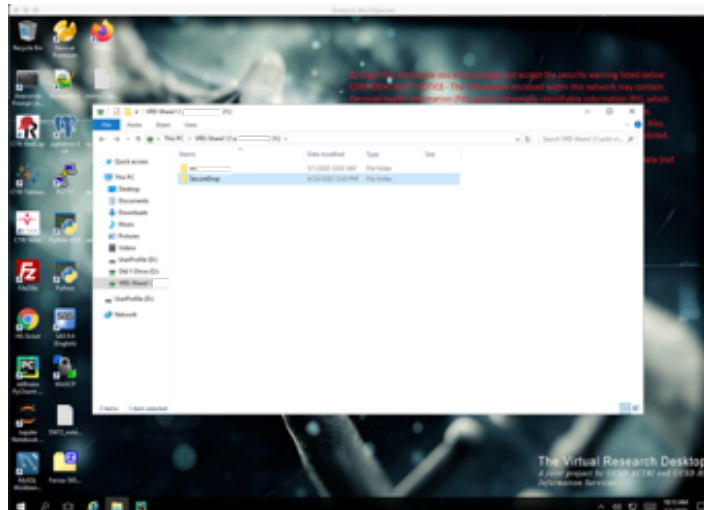
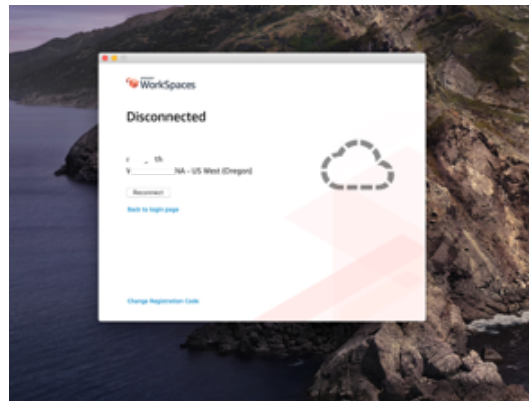
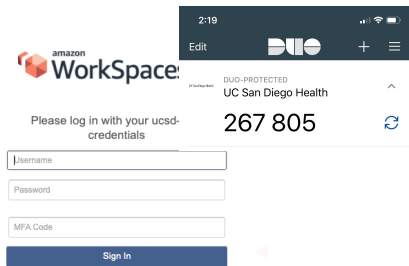
THE VIRTUAL RESEARCH DESKTOP (VRD)

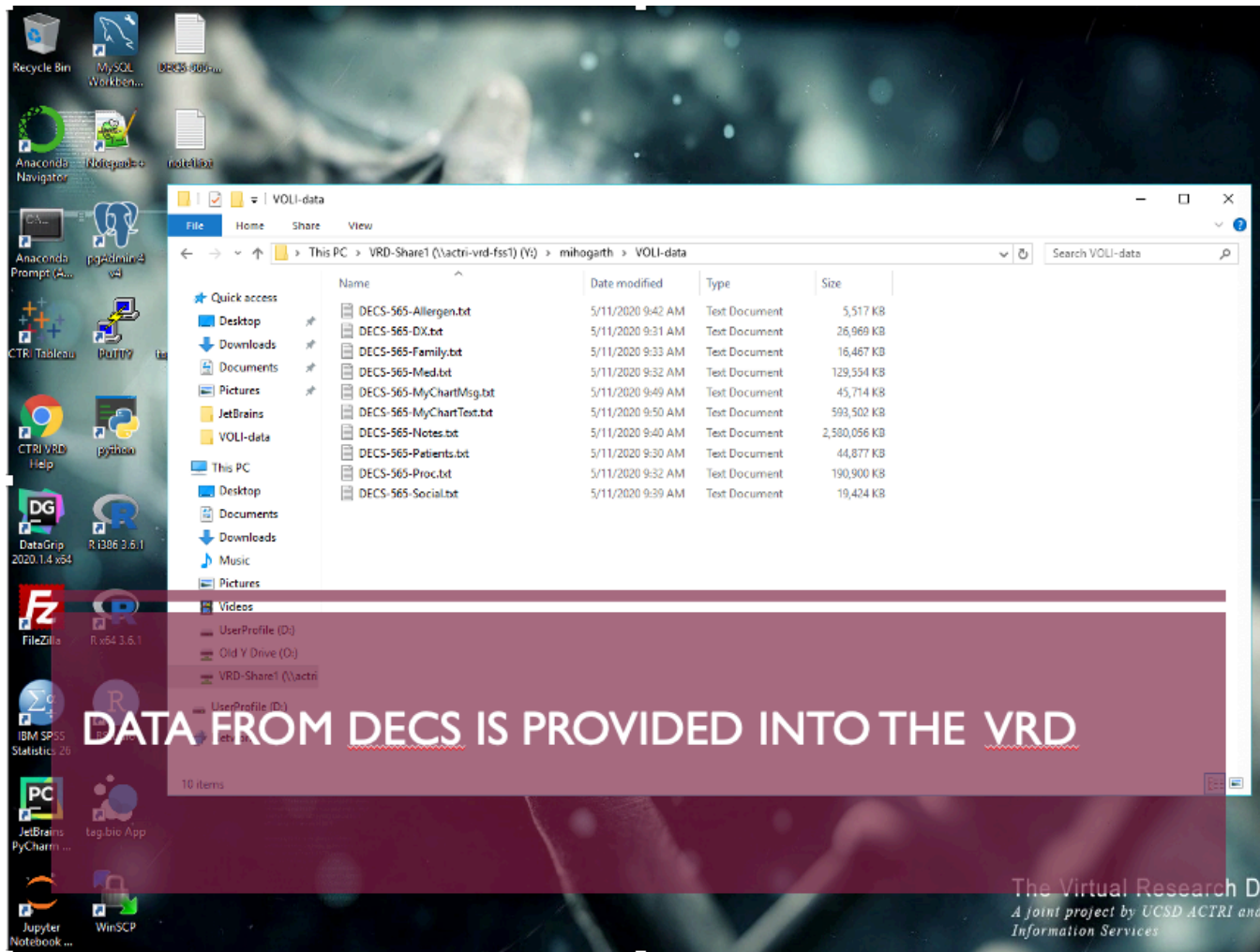
- It is a modified version of the Amazon Web Services (AWS) Windows 10 “Workspace” virtual machine
- Runs in the protected UCSDH Secure Cloud in AWS
 - in the AWS HIPAA environment
 - approved by UCSDH CSO for PHI
- Provisioned with:
 - SPSS
 - R/RStudio
 - Python/PyCharm
 - Java 8 JDK
 - Depending on approval, access to internal databases – ie, UC CORDS
 - tag.bio based access to available databases

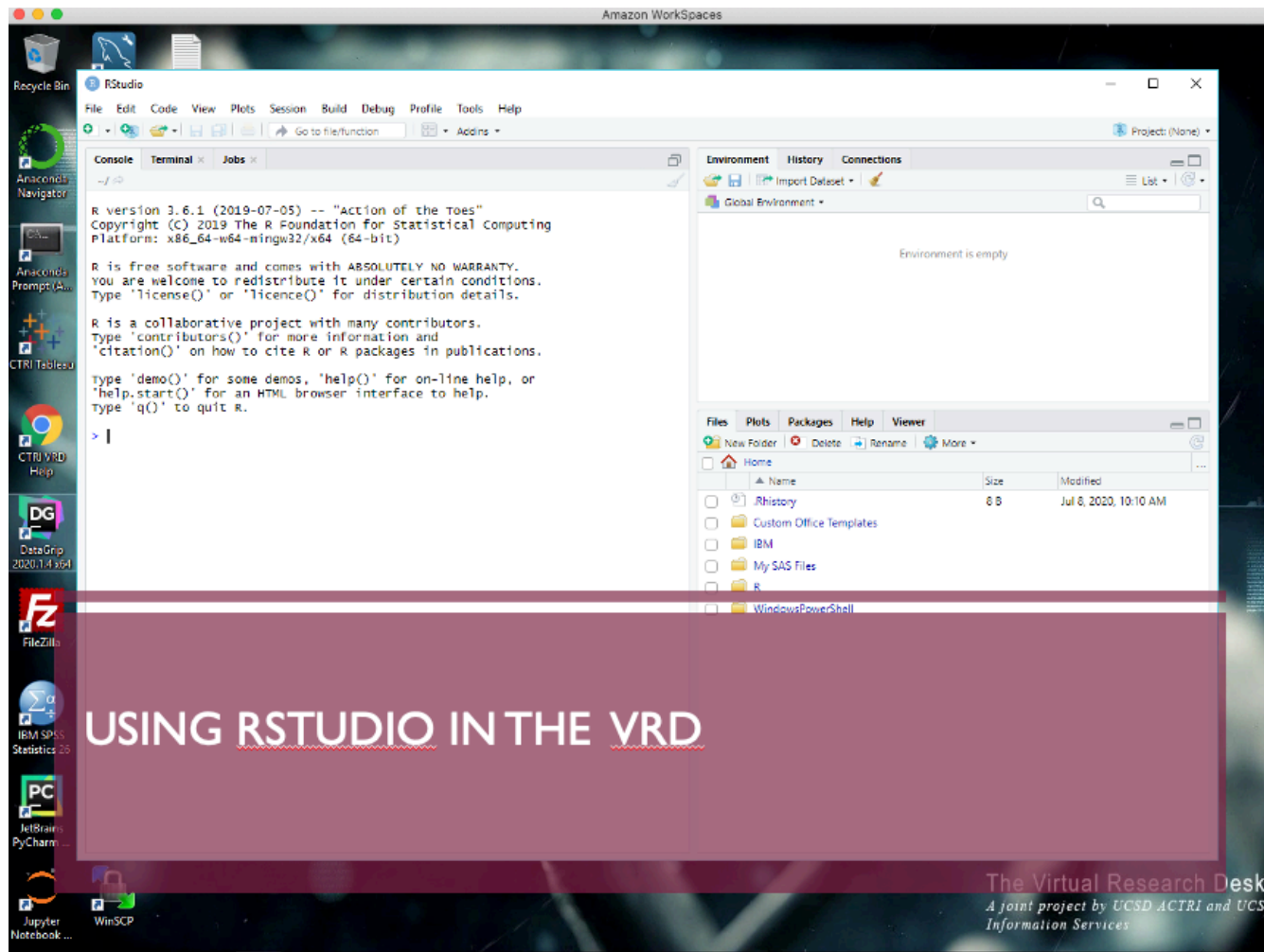


UCSD Health Virtual Research Desktop

Secure AWS workspaces: Virtual Research Desktop (VRD)







Amazon WorkSpaces

Database Consoles | CORDS-AWS | console_4 [CORDS-AWS]

Database

- AWS-MIMIC3 (1 of 7)
 - schemas 1
 - collations 222
 - users 11
- CORDS-AWS (4 of 14)
 - UCCORDS (4 of 14)
 - dbo
 - guest
 - INFORMATION_SCHEMA
 - OMOP5

```
select distinct count(*) from OMOP5.person
```

Services

- CORDS-AWS
 - console_4 5 m 15 s 664 ms
 - console_4 5 m 15 s 664 ms

Output | count(*) | int

0 rows

Comma-separated (CSV)

External file changes sync may be slow
Project files cannot be watched (are they under network mount?)

Event Log

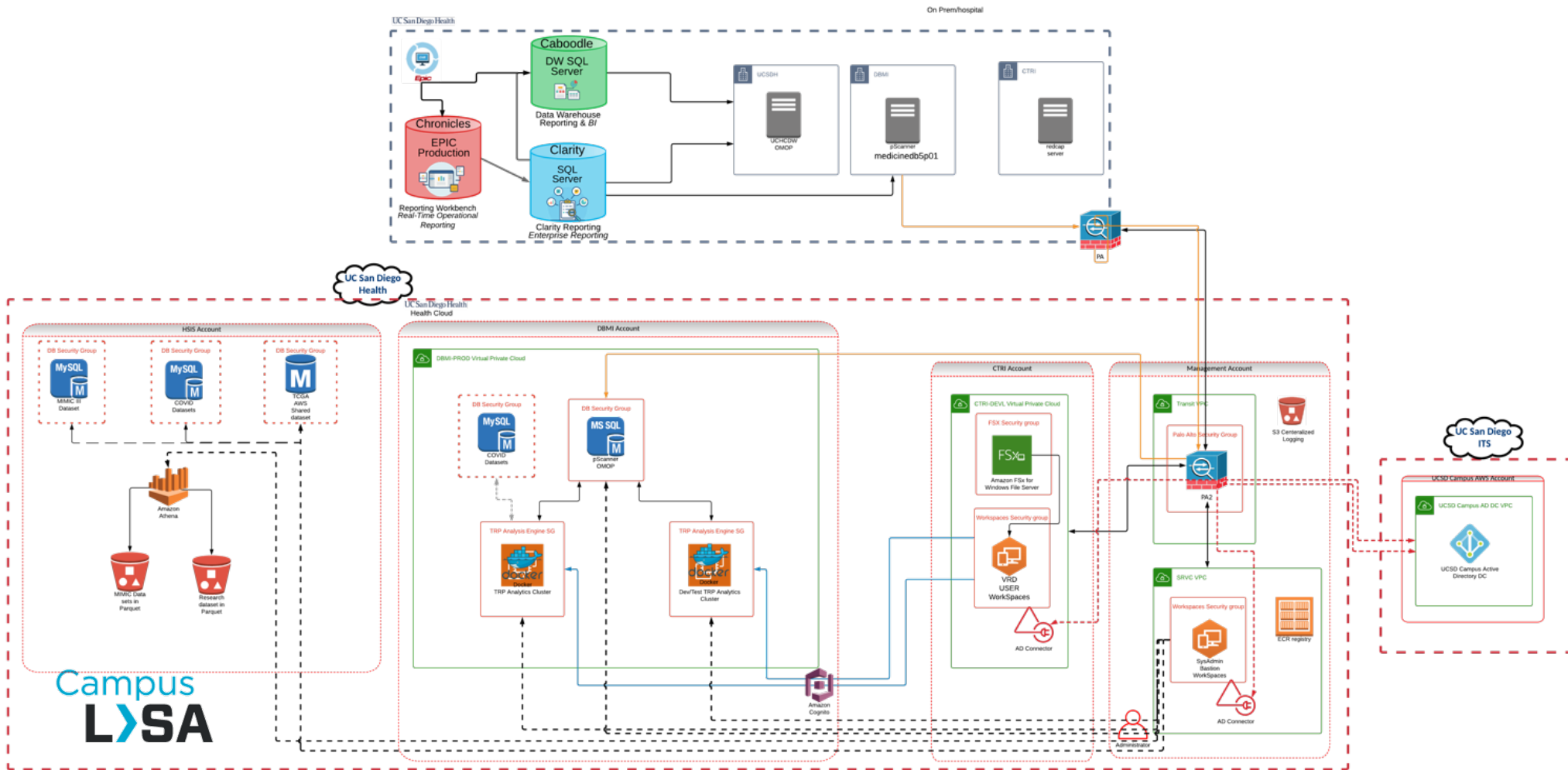
console_4 [CORDS-AWS]: select distinct count(*) from OMOP5.person... completed. (a minute ago)

1:17 UTF-8 4 spaces

The Virtual Research Desk
A joint project by UCSD ACTRI and UCSF
Information Services

ACCESSING UCCORDS IN THE UCSD SECURE RESEARCH DATA COMMONS

VRD and Multi-User Datasets



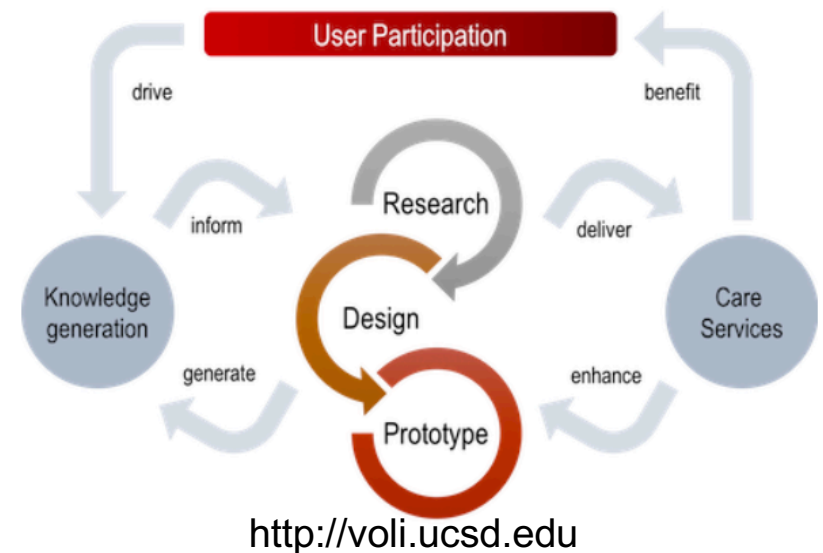
Use Case 1: VOLI Collaboration

Personalized and context-aware voice-based digital assistant to improve the quality of life and the healthcare of older adults, and consequently, to reduce caregiving burden and optimize the interactions with healthcare and service providers.



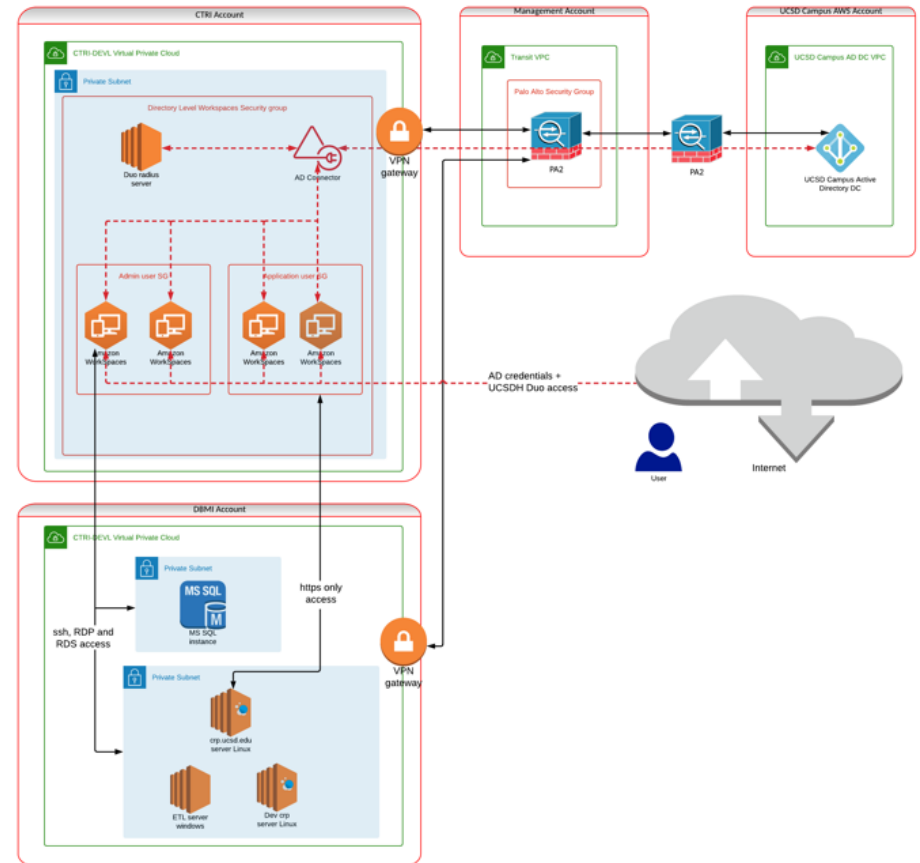
What Data is VOLI Requiring?

- Full text of all clinical notes for patients in the cohort
- All lab test data
- Patient demographics
- MyChart communications – messages between doctor and patient



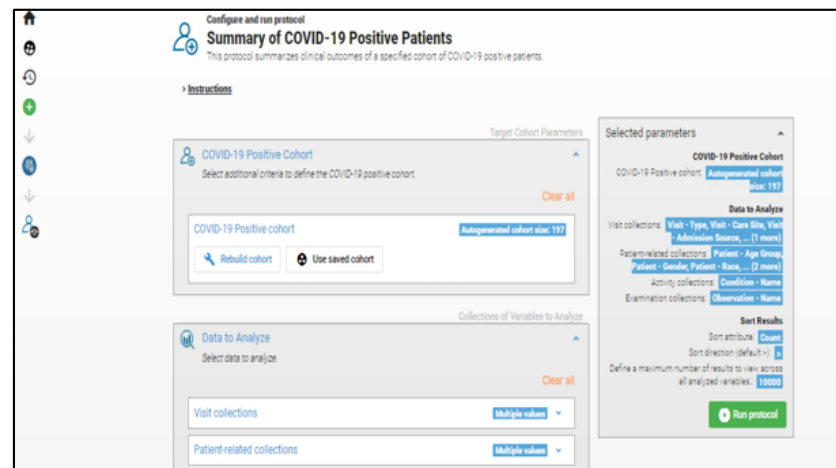
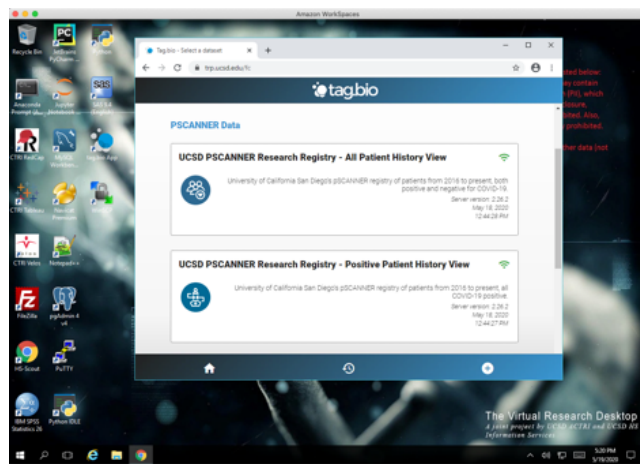
Use Case 2: Access to COVID Data – Translational Research Portal

- Analysis application for dataset exploration, building reproducible data queries
- Clinical data warehouse using a common data model (OMOP)
- This is now being used for COVID research and registry work
- Access through “Virtual Research Desktop” -- AWS Workspace confined to the research enclave

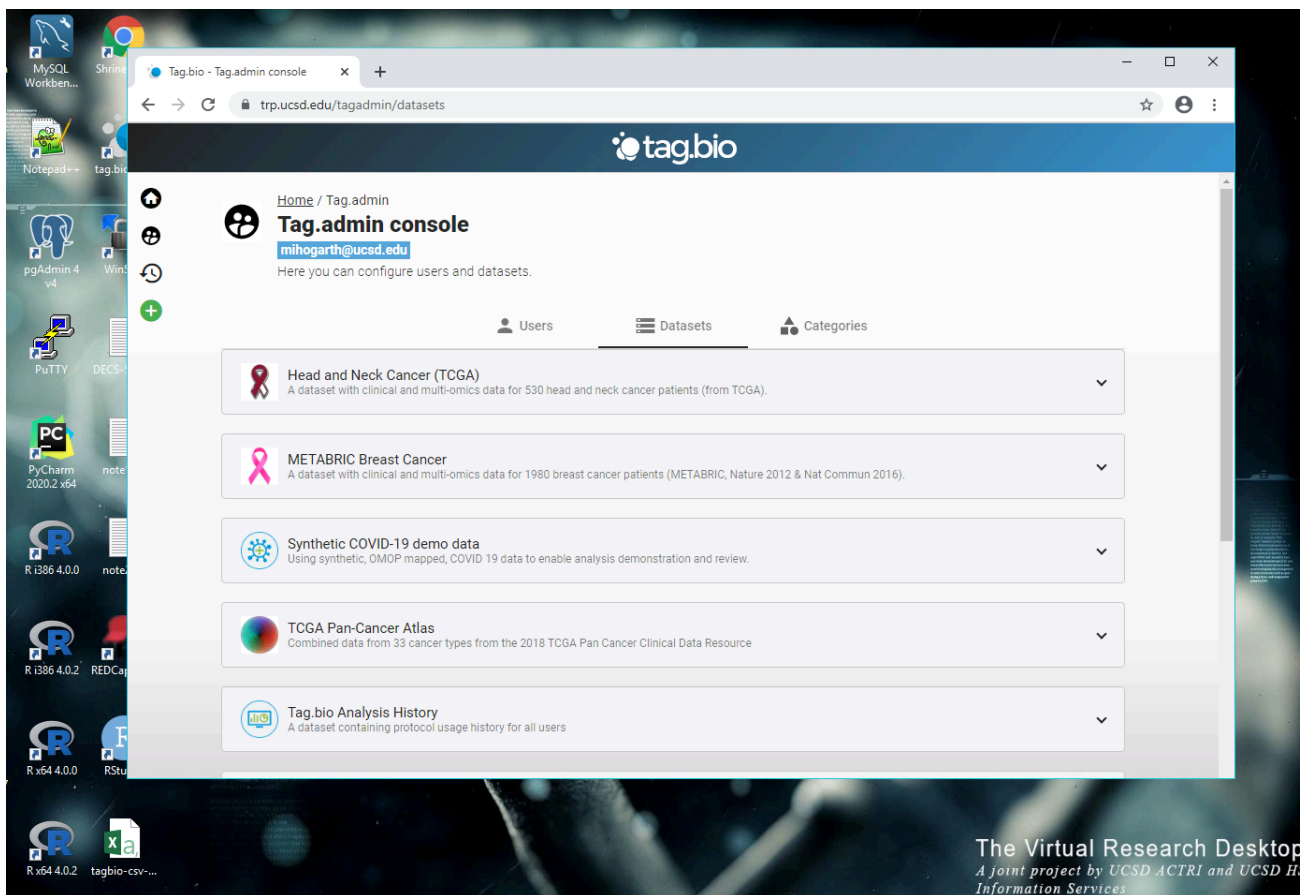


THE TRANSLATIONAL RESEARCH PORTAL: A TOOL FOR DATA EXPLORATION AND ANALYSIS

- we have installed the tag.bio system in our research cloud and it has access to data sets in our 'secure data commons database'
- the tag.bio system provides population level access and ability to perform analysis
- a user can 'slice' the cohort and select specific analyses (demographic, survival, comparison between cohorts)
- planned → with approval, provide 'download' of limited data set (LDS) row-level data from selected data set into the investigator's virtual research desktop for further analysis



Use Case 2: Access to COVID Data – Translational Research Portal Additional Dataset

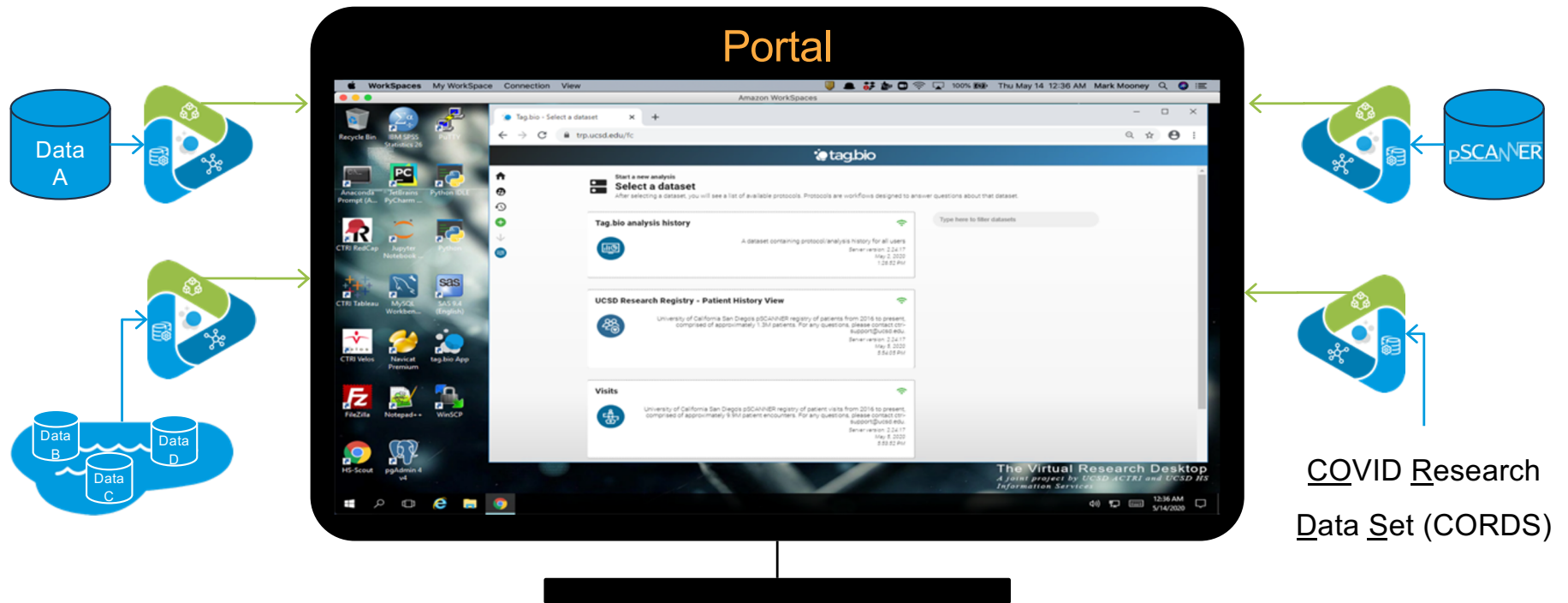


The screenshot shows a web browser window displaying the Tag.bio admin console. The browser address bar shows the URL `trp.ucsd.edu/tagadmin/datasets`. The page header includes the Tag.bio logo and navigation links for Home, Tag.admin, and the user profile `mihogarth@ucsd.edu`. Below the header, there are tabs for Users, Datasets, and Categories. The Datasets tab is active, showing a list of datasets:

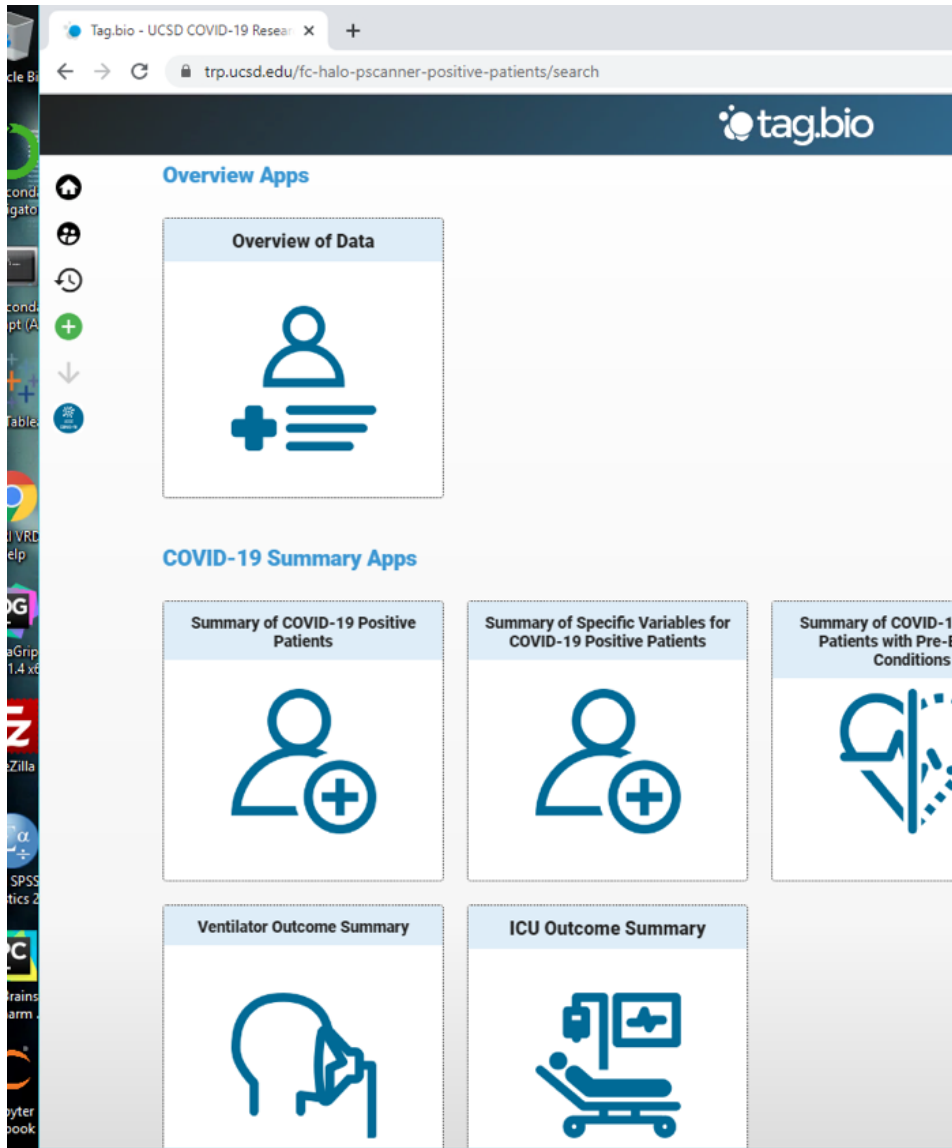
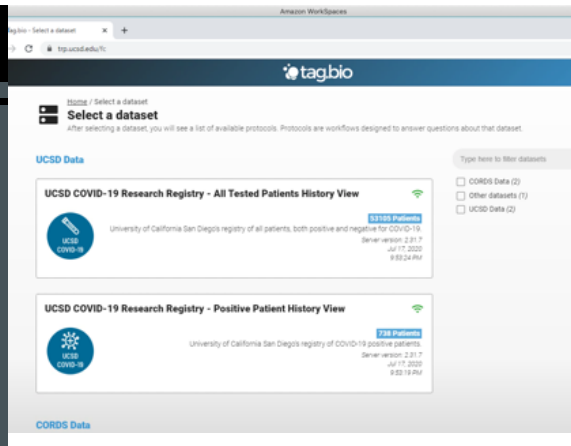
- Head and Neck Cancer (TCGA)**: A dataset with clinical and multi-omics data for 530 head and neck cancer patients (from TCGA).
- METABRIC Breast Cancer**: A dataset with clinical and multi-omics data for 1960 breast cancer patients (METABRIC, Nature 2012 & Nat Commun 2016).
- Synthetic COVID-19 demo data**: Using synthetic, OMOP mapped, COVID 19 data to enable analysis demonstration and review.
- TCGA Pan-Cancer Atlas**: Combined data from 33 cancer types from the 2018 TCGA Pan Cancer Clinical Data Resource.
- Tag.bio Analysis History**: A dataset containing protocol usage history for all users.

The desktop background shows various application icons including MySQL, Notepad++, pgAdmin 4, PuTTY, PyCharm, and RStudio. The bottom right corner of the desktop features the text: "The Virtual Research Desktop A joint project by UCSD ACTRI and UCSD HS Information Services".

The Translational Research Portal allows access to data nodes



Decentralized data. Centralized analysis.

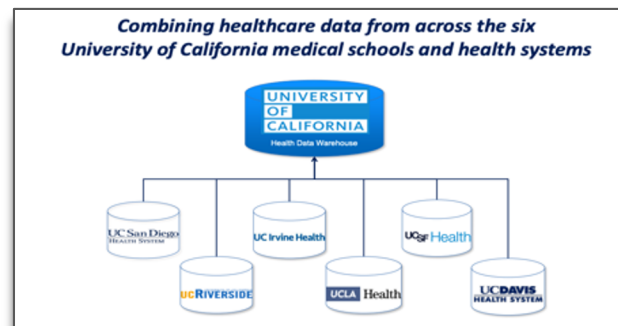
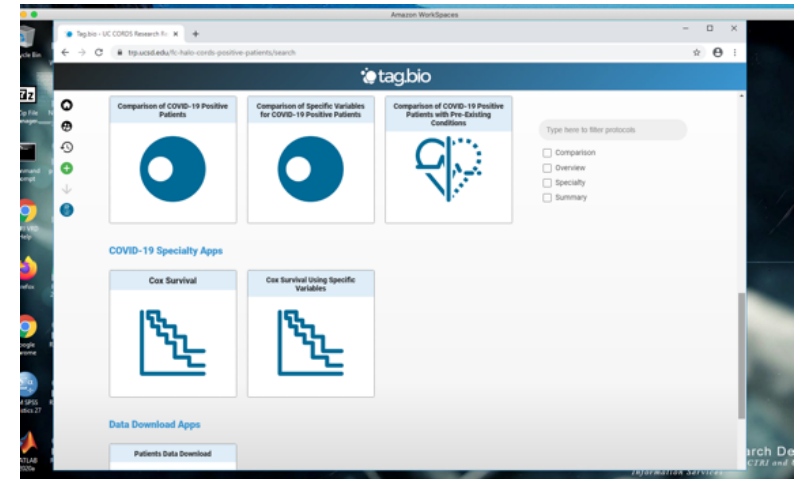


USING THE EXPLORATION TOOL

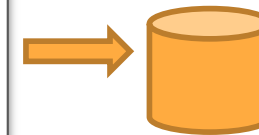
- the ucsc tag.bio system was loaded with data from the COVID-19 registry in the secure research data commons database
- the system provides a simple way to interact with the data set through “analysis protocols”

Use Case 2: UC CORDS - COVID-19

- UC Health 2019:
 - 19 health professional schools,
 - 5 academic medical centers,
 - 12 hospitals
 - 173,000 annual inpatient admissions
 - 4.8M annual outpatient visits
- UC Health Data Warehouse 2019:
 - ~5M patients seen since 2012
 - 100M encounters
 - 300M procedures
 - 1B measurements



The UC COVID Research Data Set (UC CORDS)



Aug 14 2020

- 175,517 COVID tested patients
- 6,056 COVID+ patients
- all labs, meds, vitals, 29 ICU data elements
- 319,952,837 "data points/"

Acknowledgments

- Health Information Services – Dr. Chris Longhurst, John Torello, Ken Wottge, Alan Sato, Derek Dutt, Dr. Amy Sitapati
- ACTRI – Nguyen Trieu, Perry Shipman, Quinlan Hampton, Tony Chen
- UCSD research compliance – Melissa Thrasher, Jeff Simmons, Cheryl Wagonhurst
- ITS – Declan Fleming, Nick Marangella, James Dotson, Brian DeMeulle
- Xpertech Solutions Inc. – www.Xpertech.io
- Tag.bio – Mark Mooney, Tom Covington, Jesse Paquette, Wade Webster, Kenn Broadhagen
- AWS – Dr. Prathima Srinivas, Heather Matson, Randy Ridgley, Danyell Wilt
- DBMI – Paulina Paul
- The Campus LISA Team!!!!

