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
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-PHI.

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

- **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/](https://www.cisecurity.org/cybersecurity-tools/mapping-compliance/)
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

---

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 rules
- VPC isolation
- CISO approval required
HIPAA Boundary Controls (simplified)
Multi-Researcher/Department Access Architecture
Secure AWS workspaces: Virtual Research Desktop (VRD)
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**

Investigator requests EHR extraction

DECS Analyst

Clinical Data Warehouse for Research

**data set A**

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

placed by Analyst into Investigator’s VRD – “SecureDrop” directory
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
Secure AWS workspaces: Virtual Research Desktop (VRD)
DATA FROM DECS IS PROVIDED INTO THE VRD
USING RSTUDIO IN THE VRD
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

http://voli.ucsd.edu
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 Translational Research Portal allows access to data nodes.

Decentralized data. Centralized analysis.
USING THE EXPLORATION TOOL

- The UCSD 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!!!!!!**