Navya Shah
DSCI 519

Semester Project

1. Summary of Existing Policies at The Center

This section summarizes the information protection policies which are mentioned in project documents. When creating this summary for the document information, the focus was on security relevant requirements such as access control, aspects of the CIA triad, and compliance regulations. It is organized to discuss what policies The Center has in place, then the policies Cartagenia has, then the policies of the AWS infrastructure, and lastly, an overview of NIST and HIPAA regulations.

1.1 HC-C04107 (The Center) Data Policies

A. Access Control and Data Security
    a. Access Rights: administrative side of system access, must be approved by the Bioinformatics Director or Supervisor and issued through hospital-managed LDAP.
    b. PHI used for clinical purposes unless IRB approval authorizes research use. Non-clinical users must not access PHI.
    c. Separation of duties enforced through predefined user groups:
        i. **root**: full admin control over systems, configurations, and databases. Granted to IT Director, Bioinformatics Director and designees.
        ii. **bioinfo**: access to non-clinical software, packages, configurations and databases
        iii. **bioinfoclin:** access to clinical data, can access+modify clinical software, packages, configurations and databases
        iv. **clinical**: access to clinical PHI and clinical interpretation tools. Given ability to modify pipelines or system configurations.
        v. **research**: access only to research datasets, if samples are human subjects then user may need IRB approval.
        vi. **hla**: access to clinical or research data pertaining to HLA lab only. Can access research data but if samples are humans then user may need IRB approval.

B. Logging: Usernames and access rights recorded in Computing Equipment Access Log.

C. Operational Policies
    a. Maintenance actions, such as OS updates, hardware changes etc, must be logged.
    b. Bioinformatics Director oversees data security. Lab Manager does CAP/CLIA/HIPAA compliance.
    c. Only some IT/Bioinformatics personnel may hold root access.

D. HIPAA compliance: to protect PHI under regulatory compliance
    a. Hospital hardware and clinical datasets must remain segregated
    b. All users accessing clinical data must complete required HIPAA training.
    c. All communication involving patient samples must follow hospital HIPAA policy.

E. Data Storage, Transfers and Recovery
    a. Clinical and research data stored separately with permissions for user groups.
    b. External transfers (FASTQ/VCF) require UUID replacement, encryption and checksums.
    c. PHI stored externally (AWS, USC HPCC) must reside in HIPAA-compliant environments under a valid BAA.

d. Recovery is performed by the Bioinformatics Director, with access logged
e. Data Retention
   i. Temporary: Deleted immediately after the wet-lab process per SOPs.
   ii. Short-term: 30 days post-processing.
   iii. Intermediate: 1 year post-processing.
   iv. Long-term: Until age 21 or 7 years after last treatment (whichever longer)
   v. Indefinite: Longer than long-term, no defined maximum.

1.2 Cartagenia Data Security & Confidentiality Policy (Informatics Platform)

A. Data Ownership:
   a. Cartagenia serves as a data processor without having any ownership of the data
   b. The Center retains full ownership of PHI
   c. When the contract is terminated, PHI information will be returned to the customer and removed from the active storage but backups may exist in offline storage.

B. Confidentiality:
   a. Employees and subcontractors are bound by confidentiality agreements.
   b. Cartagenia applies the same security standards to customer PHI as it uses internally.

C. Data Access Policies:
   a. Cartagenia staff may only access PHI with customer authorization or when required for time bound support tasks.
   b. All access is purpose-specific and must be documented.

D. Data Encryption:
   a. All PHI put in Cartagenia’s front-end tools uses SSL/TLS encryption.
   b. Data is stored in a dedicated database instance specific to each customer.

E. Audits and Support:
   a. Operational events are logged to support HIPAA and contractual audits.
   b. PHI used during bulk operations must be deleted or returned upon completion.

1.3 AWS HIPAA Compliance Policies (Cloud Provider)

AWS document that outlines how its cloud provider services provide HIPAA compliance

A. Encryption and Protection of PHI:
   a. Using TLS/SSH/IPsec (in transit) and KMS-backed encryption (at rest)
   b. AWS KMS (Key Management Service): to generate and store keys securely

B. Different AWS Services for PHI
   a. EC2: encrypted volumes, dedicated instances for sensitive workloads.
   b. S3: encrypted object storage with access controls.
   c. Glacier: encrypted long-term archival storage (used for NGS archives)

C. Network Security:
   a. VPC isolation ensures PHI workloads run in segregated networks.
   b. Security groups and NACLs prevent unauthorized access.

D. Audits, Storage and Recovery:
   a. CloudTrail records all API calls related to PHI
   b. S3 access logs track object-level activity
   c. VPC Flow Logs for network visibility
   d. Redundant storage (S3/EBS) supports availability and recoverability.

1.4 AWS Disaster Recovery Policies (Cloud Provider)

AWS Whitepaper that outlines how its infrastructure supports availability and continuity which replaces the need for The Center to have separate DR infrastructure

A. Recovery Objectives (RTO/RPO)
    a. RTO defines how quickly a service must be restored after failing.
    b. RPO defines how much data loss is tolerable.

B. Backup and Restore
    a. Backup Storage Options
        i. S3: durable object storage for recoverable copies.
        ii. Glacier: encrypted, long term archives (The Center's main NGS archive).
    b. Restore through retrieving from Glacier to S3 to local storage.

C. Audit and Recovery similar to HIPAA whitepaper
    a. CloudWatch and SNS notify administrators of anomalies
    b. Gameday testing which is routine validation to confirm that backups can be restored and recovery objectives are met.

1.5 NIST800-66 Security Rule Guidance

A. Implement administrative, technical, and physical safeguards for PHI.
B. Enforce least-privilege access and workforce authorization requirements.
C. Maintain audit logs for all PHI access and modification events.
D. Encrypt PHI in storage and transmission.
E. Maintain integrity protections against unauthorized alteration or destruction.
F. Train workforce members on HIPAA security requirements.
G. Maintain contingency planning (backups, DR, emergency procedures).
H. Conduct ongoing risk assessments and update controls accordingly.

1.6 PHI Identification Requirements

A. Data containing HIPAA identifiers (e.g., names, dates, addresses, MRNs) is PHI.
B. PHI created when genomic data is linked to identifiers or sample metadata.
C. Identifiers must be removed or replaced prior to external release unless IRB-approved.
D. Research personnel may not access PHI without authorization.
E. Maintain clear and documented classification of PHI vs non-PHI datasets

1.7 GUID/De-identification Policies

A. Replace identifiers (HCUID) with GUID/UUID before external distribution
B. Filenames and internal metadata should not expose identifiers.
C. Maintain a secure and access-controlled GUID-to-patient mapping.
D. Release deidentified data only under IRB or clinical policy conditions.
E. Use checksums to verify data integrity after deidentification
F. Document the procedures to prove HIPAA compliance

Most Important Info Security Policies:

1. Access Control
    a. Enforce role-based access limits and restrict PHI to clinical or IRB-approved users (HC).
    b. Limit vendor access to PHI to explicitly authorized, time-bound support tasks (Cartagenia).
    c. Control PHI access using IAM roles, bucket policies, and VPC isolation (AWS).
    d. Restrict access to PHI based on workforce authorization and least privilege (NIST).
2. Encryption
    a. Require encrypted transfer of all clinical data (SSH/SFTP/HTTPS) (HC)
    b. Enforce TLS for all platform access (Cartagenia).

    c. Encrypt PHI in transit and at rest using HIPAA-eligible cloud services and KMS (AWS).
    d. Protect PHI in storage and transmission through secure configurations (NIST).
3. Confidentiality
    a. Allow PHI use only for clinical purposes unless IRB-approved for research (HC).
    b. Maintain confidentiality agreements for all platform personnel (Cartagenia).
    c. Use only HIPAA-eligible cloud services under a signed BAA (AWS)
    d. Ensure PHI is accessed only by authorized personnel following HIPAA standards (NIST).
    e. Treat genomic data as PHI when linked to identifiers (PHI Identification).
4. Logging and Monitoring
    a. Log all system access, permissions, and recovery events (HC).
    b. Maintain audit trails for operational activity (Cartagenia)
    c. Enable CloudTrail, S3 logs and CloudWatch monitoring for PHI access and changes (AWS).
    d. Maintain auditable logs of PHI access and disclosure events (NIST).
5. Retention, Backup and Recovery
    a. Enforce retention periods and ensure archived PHI can be restored when required (HC).
    b. Archive long-term genomic data in Glacier and maintain recoverability (AWS).
    c. Maintain contingency and recovery procedures (NIST).
6. Data Segregation
    a. Keep clinical and research data strictly separated by permissions (HC).
    b. Use isolated per-customer database instances (Cartagenia).
    c. Segregate clinical vs. research datasets via IAM and S3 bucket policies (AWS).
7. PHI Identification and De-identification
    a. Remove or replace HIPAA identifiers before external data release (PHI Identification).
    b. Replace HCUID with GUID/UUID and restrict access to identifier mappings (GUID Policy).
    c. Release de-identified data only under approved clinical or IRB conditions (GUID Policy).

2. Summary of Potential Threats to PHI

Protected information, such as PHI, genomic data, clinical data, that The Center works with is vulnerable to several risks that can compromise confidentiality, integrity or availability of it.

2.1 Unauthorized Access

A. Credentials can be stolen by an attacker who can login as root/clinical.
B. There is a risk that users outside the clinical user group, like the research staff, can get access to the data they shouldn't. This can be because of misconfiguration or oversight.
C. Data that a vendor accesses can also be exposed if not managed or limited (to a time or section).
D. Cloud IAM misconfigurations or weak account security can open more doors to this.

2.2 Improper Disclosure of Identifiable Data

A. Genomic files can reveal patient identity if identified aren't removed completely or even if metadata or another file would accidentally reveal information

B. Such as mishandling the GUID/UUID mappings that can allow someone to decipher the deidentification

2.3 Loss of Important Data

A. Could be through system failures, hardware breaking, corrupted files or accidently deleting something
B. Even backup systems can fail if not tested

2.4 Data interception during transfer:

A. If PHI is moved outside secure channels or encryption is not applied correctly then attackers could intercept sensitive data in transit, especially during patient/physician file delivery

2.5 Cloud misconfiguration and third-party weaknesses:

A. Issues such as overly permissive IAM roles, misconfigured S3 buckets, expand the attack surface
B. There can also be vulnerabilities in vendor platforms like Cartagenia which can bring more risk if poorly managed

2.6 Compliance and procedural failures:

A. Gaps in de-identification, retention, logging, or HIPAA handling can convert mistakes into significant breaches, especially when dealing with PHI and long-term genomic archives.

3. Access Control Policy

This section will translate the policies from human-language to access control rules which can be enforced in an information management system.

a)
Identifying Subjects:

1. **root**: system administrators
2. **bioinfo**: non-clinical bioinformatics staff
3. **bioinfoclin**: clinical bioinformatics staff
4. **clinical**: clinicians and analysts who interpret results
5. **research**: research staff, PHI access needs IRB approval
6. **hla**: hla users (like research but for hla, unless human samples, then IRB approval)
7. **vendor**: could be Cartegenia support staff or even AWS, temporary approved access

Identifying Protected Resources:

1. **Clinical PHI datasets:** fastq files, VCF files, clinical reports, metadata of the reports
2. **Clinical configurations files**: clinical software, configurations, databases, scripts.
3. **Research datasets:** non PHI
4. **HLA clinical and research files:** clinical and research but only HLA specific
5. **GUID/UUID mapping files**: reidentification keys
6. **System logs and audit records**: maintenance logs
7. **Cloud buckets:** S3 clinical, S3 research, Glacier archives
8. **Cartagena database instance** (for each customer)
9. **Local servers/storage**: different hardware for clinical and research data storage
10. **Backups**: done for server and storage through TSM to USC HPCCso wouldnt

b)
Applying DAC and MAC to Implement Access Policies: The Center's policies need both mechanisms for its enforcement. This is because DAC is better suited for day-to-day operations, however the regulatory requirements with PHI mean that we have to add a MAC layer to protect its confidentiality and integrity.

3.1 DAC: Roles, ACLs, and Capabilities

To implement DAC, existing user groups are treated as roles and ACLs are used for each object class

- Roles: root, bioinfo, bioinfoclin, clinical, research, hla, vendor
- Objects: clinical PHI, clinical configs, research datasets, hla data, GUID mapping, logs, cloud buckets, Cartagenia database, and backups

Object / Role	**root**	**bioinfo**	**bioinfoclin**	**clinical**	**research**	**hla**	**vendor**
Clinical PHI	RW	-	R	R	- (IRB)	R (HLA clinical only)	Cap (time-limited)
Clinical Pipeline Configs	RW	R/W (non-clinical only)	RW	RW	-	R/W (HLA pipelines only)	-
Research Data	RW	RW	R	-	RW	R/W (HLA research only, IRB)	-
HLA Data	RW	-	-	-	-	R/W	Cap (time-limited)
GUID Mapping	RW	-	RW	R	-	-	-
System Logs	RW	-	-	-	-	-	-
S3 Clinical Bucket	RW	-	RW	R	-	R (HLA part only)	-
S3 Research Bucket	RW	RW	R	-	RW	R/W (HLA part only)	-
Glacier Archive	RW	-	Restore/RW	Request restore	-	Request (HLA only)	-

Explaining the DAC ACLs and Capabilities:

1. Clinical PHI:
    a. **root** -> RW, full access for admin and system level tasks
    b. **bioinfo** -> no access because its PHI
    c. **bioinfoclin** -> R, must access to clinical data to maintain clinical software/configs
    d. **clinical** -> R, because they need to work with the clinical data

e. **research** -> no access but capability from IRB for human samples can be given
f. **hla** -> R, for hla-specific data only
g. **vendor** -> capability only, can be temporary access for support

2. Clinical pipeline:
    a. **root** -> RW, full access for admin and system level tasks
    b. **bioinfo** -> R, only non-clinical pipelines
    c. **bioinfoclin** -> RW, authorized to modify clinical pipelines
    d. **clinical** -> RW, authorized to modify clinical pipelines
    e. **research** -> no access (IRB gives Capability)
    f. **hla** -> R, for hla-specific pipelines only
    g. **vendor** -> capability only, can be temporary access for support
3. Research data:
    a. **root** -> RW, full access for admin and system level tasks
    b. **bioinfo** -> RW, need full access to non-clinical data
    c. **bioinfoclin** -> R, can read to verify but not overwrite it
    d. **clinical** -> R, can read to inform reporting but not work on it
    e. **research** -> no access (IRB gives Capability)
    f. **hla** -> RW, for research in the hla domain (IRB for PHI)
    g. **vendor** -> customer controlled so no access
4. HLA data:
    a. **root** -> RW, full access for admin and system level tasks
    b. **hla** -> RW full access
    c. **vendor** -> capability only for maintenance
    d. **others** -> can be granted capabilities
5. GUID mapping
    a. **root** -> RW, full access for admin and system level tasks
    b. **bioinfoclin** -> RW, required to link pipeline output to identifiers
    c. **clinical** -> R, read needed for interpretation
    d. **others** -> no access as most sensitive
6. System logs: **root** -> RW, full access for admin and system level tasks
7. S3 Clinical bucket:
    a. **root** -> RW, full access for admin and system level tasks
    b. **bioinfo** -> no access
    c. **bioinfoclin** -> RW, manages clinical files and pipeline files
    d. **clinical** -> RW, can read the PHI stored but only modify pipeline files
    e. **research** -> no access (IRB gives Capability)
    f. **hla** -> R (only hla files)
    g. **vendor** -> no access to PHI buckets
8. S3 Research bucket:
    a. **root** -> RW, full access for admin and system level tasks
    b. **bioinfo** -> RW, manages research pipelines
    c. **bioinfoclin** -> R, may review but no modifying
    d. **clinical** -> R, may review but no modifying
    e. **research** -> RW, full access to research data
    f. **hla** -> RW only for hla research
    g. **vendor** -> no access
9. Glacier archive:
    a. **root** -> RW, full access for admin and system level tasks
    b. **bioinfoclin** -> Restore/RW, to initiate restore of data or write corrected files
    c. **clinical** -> Request restore only, can ask for older data
    d. **hla** -> Request restore only, can ask for older data (hla specific)

3.2 MAC: Levels, Clearances, Categories

The MAC layer is put on top of DAC to prevent bypassing confidentiality rules through misconfigured ACLs or mistakes.

Confidentiality levels (BLP)

1. C0: non-sensitive outputs which would be publicly disclosed
2. C1: deidentified and research data
3. C2: Clinical PHI such as NGS data, reports or genomic data
4. C3: GUID/UUID mapping used for reidentification

C3 > C2 > C1 > C0

Categories:

1. CL: clinical workflows
2. R: research workflows
3. H: hla workflows

Subject Clearances:

1. root: (C3, {CL, R, H}) to get access to everything
2. bioinfoclin: (C3, {CL}) to access both PHI and mapping tables
3. clinical: (C2, {CL}) to access PHI but not mapping tables
4. bioinfo, research: (C1, {R}) to access research data only
5. hla: (C2, {H}) to access hla data only
6. vendor: (C0, {R}) to access the least and more if only authorized

Confidentiality rules:

1. No read up
    a. Research roles (C1, {R}) cannot read any clinical PHI (C2, {CL}) or any HLA resource (C2/C1, {HLA})
    b. Clinical users (C2, {CL}) can't read GUID tables (C3, {CL}) or any HLA data (C2/C1, {HLA})
    c. Bioinfoclin (C3, {CL}) cannot read HLA data (C2/C1, {HLA})
2. No write down
    a. HLA users (C2, {HLA}) cannot write its PHI into general clinical objects (C2, {CL}) or research spaces (C1, {R})
    b. Clinical users (C2, {CL}) cannot write PHI into research buckets (C1, {R}) or into HLA categories
    c. Bioinfoclin (C3, {CL}) cannot put PHI, pipeline outputs, or identifiers into research areas (C1, {R}) or into HLA categories
3. Root bypasses all confidentiality constraints for system administration and recovery tasks.

Integrity levels (Biba)

1. I0: lowest trust inputs, external uploads
2. I1: research datasets
3. I2: clinical PHI
4. I3: clinical pipelines and configs

I3 > I2 > I1 > I0

Subject integrity levels:

1. root, bioinfoclin: I3, responsible for maintaining system/pipelines
2. Clinical: I2, uses pipelines and its results
3. hla: I2-I3, depends on workflow
4. bioinfo, research: I1, mostly working in research
5. Vendor: I0, can increase on context

Object Integrity labels:

1. I3: Clinical pipelines, HLA pipelines, System configs
2. I2: Clinical phi, HLA phi
3. I1: Research
4. I0: Vendor logs or temporary files

Integrity rules:

1. No read down:
    a. High-integrity roles (I3: root, bioinfoclin) don't read lower integrity research files (I1) to avoid contamination.
2. HLA maintainers (I3 within {HLA}) do not read general research datasets (I1, {R}) unless the data is needed in HLA workflow.
3. No write up:
    a. Research users (I1) cannot write to clinical pipelines (I3) or to clinical PHI (I2)
    b. Clinical users (I2) cannot modify clinical or HLA pipelines (I3)
    c. Bioinfo (I1) cannot modify any object above I1, including clinical data (I2)
    d. vendor (I0) cannot write to anything other than temporary diagnostic spaces.
4. Pipeline configs and GUID tables are always treated as I3, which means only root and designated pipeline maintainers write to them

Assumptions:

1. The document says clinical users can modify the pipelines but I interpreted it as more modifying clinical software, but the bioinfoclin would be the main group responsible for editing pipeline code and configs.
2. Clinical users do not need to see research or HLA data since they work limited to the clinical domain, so MAC compartments enforce this separation.
3. Bioinfo is strictly non-clinical as they work with research tools and data.
4. Only hla users (and root) can access HLA data or pipelines.
5. IRB approval is treated as a capability that temporarily grants access when needed.
6. Vendors are given capabilities for time-bound access only when needed

4. Additional Requirements Not Covered by Access Control

Beyond confidentiality and integrity, there are security and operational requirements that can be outside the scope of MAC and DAC. These gaps come from NIST800-66, HIPAA security rule and The Center's workflow.

Availability requirements:

1. System availability: clinical pipelines must be accessible for patient care
2. Backup integrity: while we do classify the access to these, there is no guarantee that these would even exist under DAC and MAC

3. Recovery: making sure it's timely and a plan is in place.

Operational requirements

1. Maintenance procedures such as patches, updates to the OS
2. SDLC security: validating the pipeline versions
3. Audit review: must review the logs created by access control
    a. Also about how permissions/roles change beyond just who accesses

Admin/Compliance requirements:

1. HIPAA training for the workplace
2. IRB approvals and tracking when its granted/expires
3. Incident response and recovery like breach investigation workflows

Physical safeguards:

1. Must protect against physical threats to workstations or servers like accessibility to it
2. Disasters that can impact the infrastructure of the hospital (outage)
3. Having a chain of custody for devices storing PHI

5. Threats Addressed By Access Control Policies And Which Are Not

Using DAC and MAC help mitigate:

1. **Unauthorized internal access:** This means that access control would help prevent unauthorized staff from accessing PHI or GUID mapping files which are more sensitive. Such as research can't access clinical PHI and clinical can't access the mapping. Also HLA data is restricted to that domain and hla users only.
2. **Misuse of PHI by legitimate users:** MAC imposes the no read up and no write down which stops data leakage into lower sensitivity buckets. This can also prevent corruption of files across the categories like clinical, research and hla.
3. **Cloud misconfigurations leading to data exposure**: Having S3 segmentation in the DAC section helps prevent the misconfigurations as the access is mapped out. The MAC categories imposed also prevent information leaking across domains (clinical into research)
4. **Modification of pipelines that is not allowed**: only high integrity users can modify
5. **Reidentification risk:** To prevent this, the GUID mapping table is kept at highest integrity and confidentiality

Threats not fully addressed:

1. **Credential stealing and/or phishing**
2. **Data loss** due to backup failures
3. **Cloud outages**: Regional failures, so no availability or business continuity at that time
4. **Mistakes when doing deidentification**: Having DAC and MAC labels in place can't ensure that the identifiers are correctly removed
5. **Insider threat**: Having a valid user with bad intentions can cause damage within their scope
6. **Network level attacks**: This could mean a man in the middle attack or eavesdropping. This would require encryption and network controls, not access controls.
7. **Physical threats**: Devices being stolen, access being compromised in the hardware servers

8. **Vendor platform vulnerabilities**: ACLs do not mitigate the threats and vulnerabilities present in Cartagenia systems or AWS infrastructure.

6. Recommended Additional Controls

To close the gaps that can't be covered by only using access control, The Center should make use of administrative, technical and physical safeguards which align with HIPAA and NIST 800-66.

Administrative controls:

1. **Workplace training and authorization:** The Center should implement annual HIPAA and security training for all the staff. It can also include more specialized training depending on your access control classifications and user groups, such as bioinfoclin being trained in maintaining security in the software pipelines. For the IRB approval, there can also be a system to track approvals (give authorization) for research users to access human subjects data. Lastly, there must be procedures in place for on/offboarding access when user groups change.
2. **Audit and monitoring:** The Center can enforce daily auto reviews of the collected CloudTrail and S3 access logs provided by AWS. There can also be an audit review taking place on a weekly basis by the Bioinformatics director. Having role reviews every quarter is also important to make sure that every user is accessing what they should be.
3. **Vendor and third party governance:** There should be a strict enforcement of having time bound access with the capabilities expiring for vendors and a way to keep a track of that. To be BAA compliant, you can conduct a security review of your vendors annually.

Technical controls:

1. **Multi-factor authentication:** Having this in place would help abate credential theft or unauthorized access. This can be placed as a requirement for all clinical, bioinfoclin, and admin accounts. It can also be required on the third-party side where it can be used for AWS IAM or Cartagenia account access, so it is confirmed the right person signed in.
2. **Network Security Enhancements:** This can be done by enforcing TLS for all data transfers, ensuring secure communications and having encryption for data in transit. The Center can also restrict outbound traffic from clinical hosts to only known destinations like a whitelist. There can also be Private VPC endpoints for S3 or Glacier that would avoid public internet use altogether.
3. **Endpoint hardening:** Important to have disk encryption on all clinical and research workstations. The Center can also require having auto-lock and inactivity timeout settings. Lastly, they can deploy endpoint detection and response (EDR) which would help pick up on malware and credential thefts on different devices on the network.

Physical controls:

1. Having guards and checkpoints in place
2. Badge controlled access to labs and server rooms
3. Having a secure storage for removable drives with a chain of custody log in place
4. Having clinical areas designed in such a way to avoid shoulder-surfing so credentials not stolen easily
5. Surveillance equipment in place like cameras in sensitive information or server rooms.

Final Summary

This report began with outlining the different documents and how The Center maintains a strong foundation in information protection policies through its internal HC-C04107 standards, Cartagenia's practices for data privacy, and AWS' HIPAA compliant configurations. The Center employed clear separation of duties through its use of role based access controls by separating users in rules such as clinical, bioinfoclin, bioinfo, research, and HLA, ensuring that PHI and clinical pipelines are accessed only by authorized groups. There are also additional safeguards in place such as encryption (at rest and in transit), deidentification of PHI, audit logging and third-party access restrictions, which would together promote confidentiality and accountability. Having data backup policies, data retention rules, and even AWS Glacier helps support the handling of clinical data which is sensitive. All of these measures together build a strong security baseline which is focused on minimizing unauthorized access and ensuring proper data handling.

However, the threat space can expand beyond what these controls work to fix. DAC and MAC can effectively prevent unauthorized access and cross-domain data leaking, along with enforcing least privilege, there are still multiple risks that remain. To name a few, it could be credential stealing, network attacks, or cloud misconfigs that can bypass permissions in place. Insider misuse can also still be possible as the model doesn't implement as many operational requirements such as audit review, patching or IRB tracking. Even physical threats can amount to serious breaches. These gaps just highlight what is needed for HIPAA and NIST 800-66 alignment. The paper then suggests the controls for each section to help fill these gaps in security. Doing so, would extend The Center's confidentiality focused policies towards more of a comprehensive security posture that addresses availability, resilience, and the full range of threats identified in this assessment.

References (beyond the documents provided):

1. Amazon Web Services. Amazon Simple Storage Service User Guide. Amazon Web Services, https://docs.aws.amazon.com/AmazonS3/latest/userguide/GetStartedWithS3.html
2. Amazon Web Services. Amazon S3 Glacier User Guide. Amazon Web Services, https://docs.aws.amazon.com/AmazonS3/latest/userguide/security-iam.html
3. Amazon Web Services. AWS CloudTrail User Guide. Amazon Web Services, https://docs.aws.amazon.com/awscloudtrail/latest/userguide/cloudtrail-user-guide.html
4. Amazon Web Services. Access an AWS service using an interface VPC endpoint. Amazon Web Services, https://docs.aws.amazon.com/vpc/latest/privatelink/create-interface-endpoint.html