NIST SPECIAL PUBLICATION 1800-43C

# Genomic Data Threat Modeling: Privacy

## An Implementation for Genomic Data Sequencing and Analysis

**Ronald Pulivarti**
National Cybersecurity Center of Excellence
National Institute of Standards and Technology

**Justin Wagner**
Material Measurement Laboratory
National Institute of Standards and Technology

**Brett Kreider**
**Stuart S. Shapiro**
**Julie Nethery Snyder**
**Kevin E. Wilson**
**Martin Wojtyniak**
The MITRE Corporation

**Scott Ross**
**Philip Whitlow**
*HudsonAlpha Institute for Biotechnology*

**Isabelle Brown-Cantrell**
**Patrick Pape**
**Jared Sheldon**
*The University of Alabama in Huntsville*

August 2025

DRAFT

This publication is available free of charge from
https://www.nccoe.nist.gov/projects/cybersecurity-and-privacy-genomic-data

## DISCLAIMER

Certain commercial entities, equipment, products, or materials may be identified by name or company logo or other insignia in order to acknowledge their participation in this collaboration or to describe an experimental procedure or concept adequately. Such identification is not intended to imply special status or relationship with NIST or recommendation or endorsement by NIST or NCCoE; neither is it intended to imply that the entities, equipment, products, or materials are necessarily the best available for the purpose.

While NIST and the NCCoE address goals of improving management of cybersecurity and privacy risk through outreach and application of standards and best practices, it is the stakeholder’s responsibility to fully perform a risk assessment to include the current threat, vulnerabilities, likelihood of a compromise, and the impact should the threat be realized before adopting cybersecurity measures such as this recommendation.

National Institute of Standards and Technology Special Publication 1800-43, Natl. Inst. Stand. Technol. Spec. Publ. 1800-43, 56 pages, (August 2025), CODEN: NSPUE2

## NIST TECHNICAL SERIES POLICIES

Copyright, Use, and Licensing Statements

NIST Technical Series Publication Identifier Syntax

## AUTHOR ORCID IDS

Ronald Pulivarti: 0000-0002-8330-3474
Justin Wagner: 0009-0003-8903-0504
Brett Kreider: 0009-0004-1508-5876
Stuart Shapiro: 0000-0003-3676-7388
Julie Nethery Snyder: 0009-0004-6352-2831
Kevin Wilson: 0009-0008-3673-6040
Martin Wojtyniak: 0009-0005-9643-2194
Scott Ross: 0009-0002-8672-6496
Philip Whitlow: 0009-0000-7677-3825
Isabelle Brown-Cantrell: 0009-0004-8820-6448
Patrick Pape: 0009-0005-4922-4026
Jared Sheldon: 0009-0009-7909-4217

## FEEDBACK

You can view or download the draft guide at the NCCoE Genomics project page.

Comments on this publication may be submitted to: genomic_cybersecurity_nccoe@nist.gov.

Public comment period: August 5, 2025 through September 4, 2025

All comments are subject to release under the Freedom of Information Act. NIST welcomes feedback and input on any aspect of this document and additionally proposes a list of non-exhaustive questions and topics for consideration:

1. How well does the threat modeling exercise in this guide relate to existing efforts in your organization? Are there significant gaps between the sets of practices that this guide should address?
2. How do you expect this document to influence your future practices and processes?
3. What changes would you like to see to increase or improve your organization's use of this document?
4. What suggestions do you have on changing the format of the information provided?
5. Is the example provided here sufficient for your organization to identify and address genomic data threats in sequencing or data analysis? Are there changes or additional content that the authors should consider?

National Cybersecurity Center of Excellence
National Institute of Standards and Technology
100 Bureau Drive
Mailstop 2002
Gaithersburg, MD 20899
Email: nccoe@nist.gov

## NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

The National Cybersecurity Center of Excellence (NCCoE), a part of the National Institute of Standards and Technology (NIST), is a collaborative hub where industry organizations, government agencies, and academic institutions work together to address businesses' most pressing cybersecurity issues. This public-private partnership enables the creation of practical cybersecurity solutions for specific industries, as well as for broad, cross-sector technology challenges. Through consortia under Cooperative Research and Development Agreements (CRADAs), including technology partners—from Fortune 50 market leaders to smaller companies specializing in information technology security—the NCCoE applies standards and best practices to develop modular, adaptable example cybersecurity solutions using commercially available technology. The NCCoE documents these example solutions in the NIST Special Publication 1800 series, which maps capabilities to the NIST Cybersecurity Framework and details the steps needed for another entity to re-create the example solution. The NCCoE was established in 2012 by NIST in partnership with the State of Maryland and Montgomery County, Maryland.

To learn more about the NCCoE, visit https://www.nccoe.nist.gov/. To learn more about NIST, visit https://www.nist.gov.

## NIST CYBERSECURITY PRACTICE GUIDES

NIST Cybersecurity Practice Guides (Special Publication 1800 series) target specific cybersecurity challenges in the public and private sectors. They are practical, user-friendly guides that facilitate the adoption of standards-based approaches to cybersecurity. They show members of the information security community how to implement example solutions that help them align with relevant standards and best practices, and provide users with the materials lists, configuration files, and other information they need to implement a similar approach.

The documents in this series describe example implementations of cybersecurity practices that businesses and other organizations may voluntarily adopt. These documents do not describe regulations or mandatory practices, nor do they carry statutory authority.

## ABSTRACT

This paper provides an example of how to conduct genomic data threat modeling for privacy on a data processing environment, including documenting the architecture, identifying threats, applying sample interventions, and iterating the process as needed. The paper complements the earlier NIST CSWP 35, *Cybersecurity Threat Modeling the Genomic Data Sequencing Workflow.*

## KEYWORDS

*DNA sequencing, genomics, genomic data, genomic sequencing, human genome, threat modeling, threat mitigations.*

## ACKNOWLEDGMENTS

We are grateful to the following individuals for their generous contributions of expertise and time.

| Name | Organization |
|---|---|
| Dylan Gilbert | NIST (former employee, all work performed while employed) |
| Justin Zook | NIST |
| Meagan Cochran | HudsonAlpha Institute for Biotechnology |
| Cherilyn Pascoe | NIST |
| Diane Wertime | NIST |
| Gary Howarth | NIST |
| Hannah Zook | NIST |

## DOCUMENT CONVENTIONS

The terms “shall” and “shall not” indicate requirements to be followed strictly to conform to the publication and from which no deviation is permitted. The terms “should” and “should not” indicate that among several possibilities, one is recommended as particularly suitable without mentioning or excluding others, or that a certain course of action is preferred but not necessarily required, or that (in the negative form) a certain possibility or course of action is discouraged but not prohibited. The terms “may” and “need not” indicate a course of action permissible within the limits of the publication. The terms “can” and “cannot” indicate a possibility and capability, whether material, physical, or causal.

## CALL FOR PATENT CLAIMS

This public review includes a call for information on essential patent claims (claims whose use would be required for compliance with the guidance or requirements in this Information Technology Laboratory (ITL) draft publication). Such guidance and/or requirements may be directly stated in this ITL Publication or by reference to another publication. This call also includes disclosure, where known, of the existence of pending U.S. or foreign patent applications relating to this ITL draft publication and of any relevant unexpired U.S. or foreign patents.

ITL may require from the patent holder, or a party authorized to make assurances on its behalf, in written or electronic form, either:

a) assurance in the form of a general disclaimer to the effect that such party does not hold and does not currently intend holding any essential patent claim(s); or

b) assurance that a license to such essential patent claim(s) will be made available to applicants desiring to utilize the license for the purpose of complying with the guidance or requirements in this ITL draft publication either:

1. under reasonable terms and conditions that are demonstrably free of any unfair discrimination; or
2. without compensation and under reasonable terms and conditions that are demonstrably free of any unfair discrimination.

Such assurance shall indicate that the patent holder (or third party authorized to make assurances on its behalf) will include in any documents transferring ownership of patents subject to the assurance, provisions sufficient to ensure that the commitments in the assurance are binding on the transferee, and that the transferee will similarly include appropriate provisions in the event of future transfers with the goal of binding each successor-in-interest.

The assurance shall also indicate that it is intended to be binding on successors-in-interest regardless of whether such provisions are included in the relevant transfer documents.

Such statements should be addressed to: genomic_cybersecurity_nccoe@nist.gov

Summary ........ 1

1 Introduction to the Guide ........ 2

1.1 Audience and Purpose ........ 2

1.2 Scope and Use Cases ........ 2

1.3 Genomic Data Characteristics ........ 3

1.4 Privacy Landscape ........ 4

1.5 Risk Modeling ........ 5

1.6 Threat Modeling ........ 6

2 Genomic Data Threat Modeling Example ........ 7

2.1 Question 1: “What are we working on?” ........ 7

2.1.1 Context ........ 8

2.1.2 Environmental Context ........ 8

2.1.3 System Context ........ 12

2.1.4 Operational Description ........ 17

2.2 Question 2: “What could go wrong?” ........ 19

2.2.1 LINDDUN Analysis ........ 20

2.2.2 PANOPTIC Analysis ........ 24

2.2.3 Threat Validation ........ 28

2.3 Question 3: “What are we going to do about it?” ........ 30

2.3.1 Threat Prioritization ........ 31

2.3.2 Response Determination ........ 35

2.4 Question 4: “Did we do a good job?” ........ 38

2.4.1 Did We Do a Good Job Documenting the System and Its Data Actions? ........ 39

2.4.2 Did We Do a Good Job Identifying and Documenting Threats? ........ 39

2.4.3 Did We Do a Good Job Responding to the Threats? ........ 40

2.4.4 Additional Activities ........ 41

3 Conclusion ........ 43

Appendix A List of Acronyms ........ 44

Appendix B References ........ 46

Appendix C Threat Modeling Approach ........ 47

Appendix D Methodology Overview ........ 47

Appendix E System Description ........ 47
Appendix F Dataflow Analysis ........ 47
Appendix G Threat Validation and Prioritization ........ 47

# List of Figures

Figure 1. Genomic Data Sequencing Workflow ........ 3
Figure 2. Genomic Data Relationships ........ 4
Figure 3. Overview of the NIST PRAM ........ 7
Figure 4. Core Example Dataflow Diagram ........ 18

# List of Tables

Table 1. PRAM Worksheet 1, Framing Business Objectives & Organizational Privacy Governance: Task 1 Questions and Responses ........ 8
Table 2. PRAM Worksheet 1, Framing Business Objectives & Organizational Privacy Governance: Task 2 Questions and Responses ........ 10
Table 3. PRAM Worksheet 2, Assessing System Design: Organizational Contextual Factors ........ 11
Table 4. Worksheet 2, Assessing System Design: Contextual Factors for Individuals ........ 11
Table 5. PRAM Worksheet 2, Assessing System Design: System Privacy Capabilities for Clinical Use Case ........ 12
Table 6. PRAM Worksheet 2, Assessing System Design: System Privacy Capabilities for Research Use Case ........ 13
Table 7. PRAM Worksheet 2, Assessing System Design: System Contextual Factors........ 13
Table 8. PANOPTIC Contextual Mapping for Clinical Use Case ........ 14
Table 9. PANOPTIC Contextual Mapping for Research Use Case ........ 16
Table 10. PRAM Data Action Types ........ 19
Table 11. LINDDUN Per Element Threat Mapping Heuristic ........ 20
Table 12. LINDDUN Dataflow Analysis for the Core Example ........ 22
Table 13. Threat Actions Identified by the PANOPTIC Privacy Activity Mapping for the Core Example........ 25
Table 14. Attack Scenarios Relevant to the Core Example ........ 27
Table 15. Core Example Attack Validations ........ 28

Table 16. Attack Difficulty Scale ..................... 31
Table 17. Core Example Threat Characteristics ..................... 32
Table 18. Attack Feasibility and Difficulty Combination Values..................... 33
Table 19. LINDDUN Threat Weights..................... 33
Table 20. Core Example Threats in Ranked Order from Highest to Lowest Priority..................... 34
Table 21. Mapping from Single Source Profiling to SP 800-53r5 Controls ..................... 36

# Summary

In this paper, the National Institute of Standards and Technology (NIST) National Cybersecurity Center of Excellence (NCCoE) demonstrates genomic data threat modeling for sample environments involved in clinical or research genomic sequencing and data analysis. This iterative, flexible modeling approach focuses on identifying threats directly to system components and data transfers in comparison to risk modeling, which emphasizes understanding potential consequences. The process examines the characteristics and methods of potential attacks to understand how they might occur and what vulnerabilities they could exploit. This paper shows a privacy-specific implementation of a common four-step threat modeling process that can be emulated by other organizations. In each of the four questions below, "we" refers to the team performing the threat modeling.

1. Document ***"What are we working on?"*** with contextual descriptions and architecture captured using worksheets adapted from the NIST Privacy Risk Assessment Methodology (PRAM) [1] and augmented dataflow diagrams for the genomic data processing environment (Section 2.1).
2. Evaluate ***"What could go wrong?"*** by identifying genomic data threats for both the clinical and research use cases using the LINDDUN [2] and MITRE PANOPTIC [3] frameworks and documenting the results using an adapted NIST PRAM worksheet (Section 2.2).
3. Determine ***"What are we going to do about it?"*** by prioritizing the identified threats to help select initial targets for interventions leveraging the NIST Privacy Framework [4], *NIST Genomic Data Profile* [5], and Special Publication 800-53r5 [6] control catalog (Section 2.3).
4. Consider ***"Did we do a good job?"*** by reviewing the results of the threat modeling exercise and identifying potential additional activities, including further interventions or continuous monitoring (Section 2.4).

Organizations rely on genomic data processing to develop biotechnology and provide clinical diagnosis. Cybersecurity and privacy risks for genomic data are complicated by the nature of the data, which is immutable and includes kinship, health, and phenotype. Further, the genomic community constitutes a broad variety of stakeholders around the world including government, academia, and industry engaged in research, healthcare, law enforcement, and direct-to-consumer genetic testing.

This paper is part of an NCCoE SP 1800 series that was developed while engaging genomic data processing stakeholders to create practical guidelines that address related cybersecurity and privacy concerns. NIST Cybersecurity White Paper 35*, Cybersecurity Threat Modeling the Genomic Data Sequencing Workflow* [7] pairs with this paper by providing similarly targeted guidelines from a cybersecurity perspective. The NCCoE Genomic Data website provides links to the overall project, including workshops and publications. Certain appendix content, containing additional resources and detailed information, is available through NIST GitHub Pages.[1]

[1] https://github.com/usnistgov/nccoe-genomic-threat-modeling

# 1 Introduction to the Guide

This document provides an example of how to conduct genomic data threat modeling for privacy on processing environments to help identify potential threats, prioritize them, and develop potential interventions. The term *threat modeling* is used here for privacy to describe a process consistent with the cybersecurity threat modeling document [7] as both cybersecurity and privacy issues can arise in genomic data processing. The environments represent a baseline implementation with devices, processes, and tools commonly used by government, academia, and industry for processing genomic data in clinical and research contexts.

## 1.1 Audience and Purpose

This paper is intended for organizations that process genomic datasets in clinical or research contexts. Genomic data processing includes sequencing genomic material as well as storing, analyzing, transferring, and appropriate destruction of genomic data. These organizations can apply the threat modeling process to develop dataflow diagrams (DFDs), identify threats, and understand interventions. Threat modeling can be used to:

- Guide system development and assess the threat reduction value of proposed threat interventions.
- Assess proposed changes to architecture or functionality for impacts on system threat and risk posture.
- Evaluate and respond to threat environment changes, such as threat intelligence or incident.
- Develop a Privacy Framework Organizational Profile that tailors the Genomic Data Profile [5] to identify and prioritize threat-informed capabilities.
- Incorporate threats into the NIST PRAM [1] by mapping the validated threats into the standard Worksheet 3 (Prioritizing Risk) based on the associated data actions, assigning relevant Problematic Data Actions and Problems for Individuals, and using the attack feasibility and difficulty combination values (converted to a 10-point scale) as surrogates for likelihood.

## 1.2 Scope and Use Cases

This threat modeling example addresses common elements of a genomics workflow including sending physical samples to a sequencing service provider, sequencing of deoxyribonucleic acid (DNA), and receiving the resulting data from the service provider. The biotechnology sector relies on elements of this workflow for many of its products and services. This workflow includes several types of entities, including commonly a *Clinical Client/Research Partner* and a *Genomic Sequencing Service*.[2] While there are distinct differences between clinical and research contexts, they share a core workflow. In the

[2] Note that client and service as used here refer to actors and not technical architecture

example workflow, a client/partner sends a specimen[3] to the genomic sequencing service to process the sample and return digital data in the form of a genomic sequence or analytical test results. The genomic sequence serves as an input to the client/partner's bioinformatics data analysis pipelines and can be used to support patient care by the Clinical Client. For this work, the NCCoE sent a DNA reference material (*Clinical Client/Research Partner*) to a genomic center (*Genomic Sequencing Service*) to sequence the sample then transfer the data back using a widely adopted protocol to the NCCoE for secondary analysis. Figure 1 illustrates this genomic sequencing workflow but does not depict the subsequent handling of data after its initial use. Organizations may either retain or dispose of data, based on its intended purpose and the organization's data retention practices and according to patient or research subject consent.

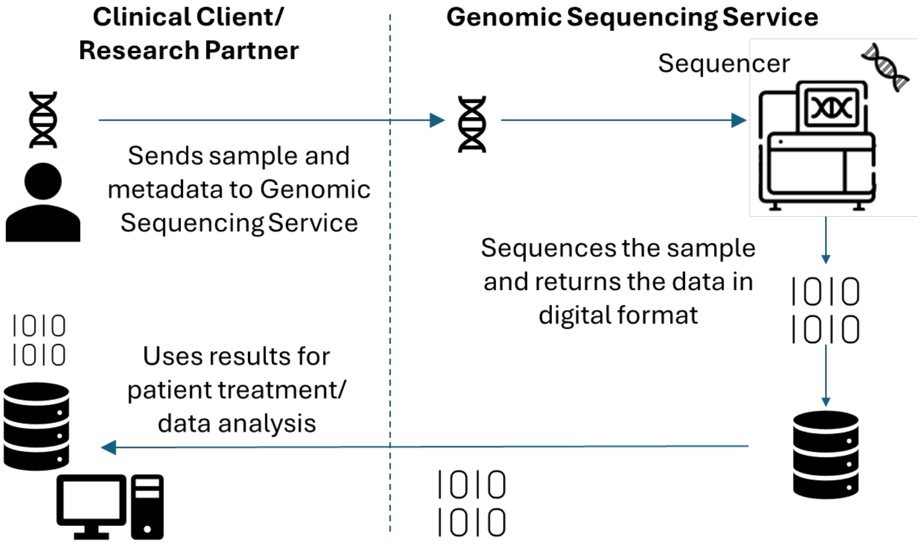

Clinical Client/
Research Partner
Genomic Sequencing Service
Sequencer
Sends sample and metadata to Genomic Sequencing Service
Sequences the sample and returns the data in digital format
1010
1010
Uses results for patient treatment/ data analysis
1010
1010
1010
1010

Figure 1. Genomic Data Sequencing Workflow

Throughout this document, we use a limited *core* example to illustrate the described methods. This *core* example is a generalized version of the analyzed processes that are common to both the clinical and the research use cases. Artifacts related to the *complete* example that are not included in the body of this paper can be found in the designated appendices. The *complete* example includes more comprehensive analysis of both the clinical and research cases.

## 1.3 Genomic Data Characteristics

The nature of human genomic data poses challenges for privacy. As a biometric it is immutable (unlike, for example, a password or a phone number). When genomic data is leaked or moves beyond a sphere

[3] Note that in some research use cases, such as re-analysis of existing data or aggregating across large sample collections, the digital genomic representation plus associated metadata may be sent to a service provider for processing

of control, the affected data subjects cannot respond by simply changing their genomes. Moreover, that durability can motivate the prolonged retention of genomic data over time, rendering it more vulnerable to eventual disclosure or misuse.

Equally problematic is the extent of the information contained in a person's genome. While the interpretability of genomic data varies, the risk to privacy extends beyond identification. Genomic data can reveal a variety of health-related conditions or susceptibility to conditions. It can also reveal family connections and in doing so imply the potential health status or predisposition of others beyond the original data subject. This is in addition to the incidental data (e.g., contact information) that these others may share with the original data subject. It is therefore useful to distinguish between direct (i.e., sample provider) and indirect data subjects (i.e., biological relatives). Figure 2 illustrates these relationships for both the clinical and research use cases, where the direct data subject is a patient and/or research subject.

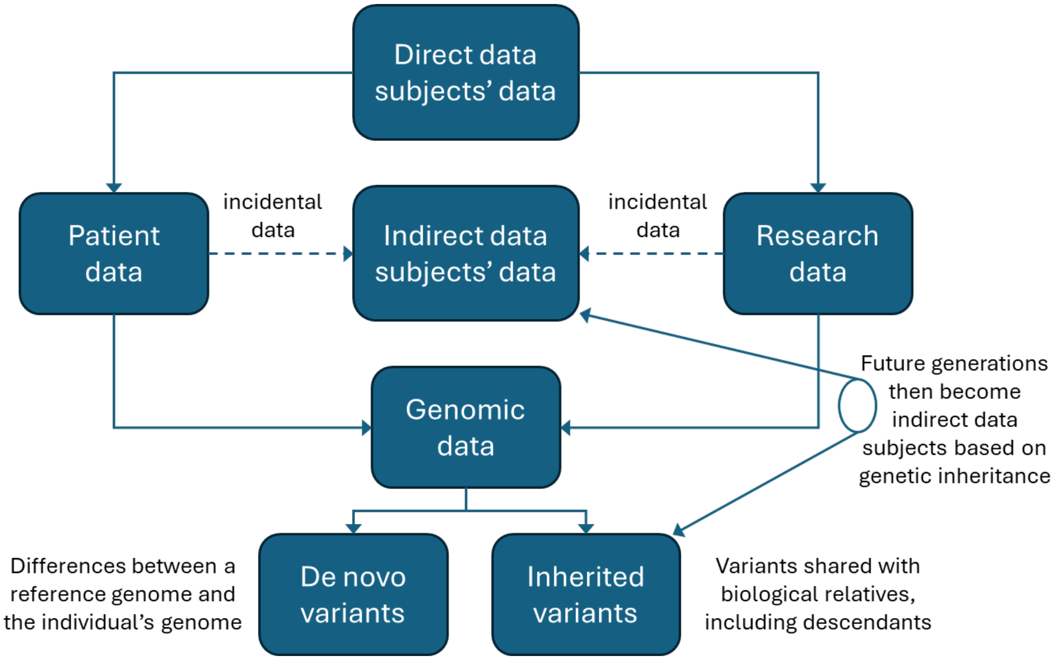

Direct data subjects' data
Patient data
incidental data
Indirect data subjects' data
incidental data
Research data
Genomic data
Future generations then become indirect data subjects based on genetic inheritance
Differences between a reference genome and the individual's genome
De novo variants
Inherited variants
Variants shared with biological relatives, including descendants

Figure 2. Genomic Data Relationships

## 1.4 Privacy Landscape

NIST SP 800-188 [8] that focuses on techniques to de-identify government datasets includes a glossary definition of privacy as, "Freedom from intrusion into the private life or affairs of an individual when that intrusion results from undue or illegal gathering and use of data about that individual," though universal agreement on a definition is still forming. However, the privacy literature includes different types of privacy associated with the contexts in which privacy problems may arise. While individual classifications may differ, they tend to resemble one another. Considering those classes specified by International Association of Privacy Professionals, of relevance to genomics are physical or bodily privacy[4] (i.e., privacy problems that deal with the human body or bodily functions) and information[5] or data privacy[6] (i.e., privacy problems that arise based on how data is processed). In the context of genomics, physical or bodily privacy applies to the acquisition of biospecimens from individuals while information or data privacy applies to symbolic representations of those specimens and any information derived from them, as well as accompanying metadata or identifiers (e.g., medical record numbers), demographics (e.g., age, gender), and diagnostic codes.

Those individuals to whom information or data pertain are often referred to as "data subjects" to emphasize the connection between the two. Information or data privacy is often confused with data security owing to their common interest in confidentiality (protecting data from unauthorized access or disclosure). However, data confidentiality is only one facet of data privacy out of many, including aspects of control over data and constraints on the collection and use of data. (While privacy is dependent on security, that dependency is not explicitly covered here given the cybersecurity threat modeling described in NIST CSWP 35 [7].) This broader landscape of privacy is recognized in systems-level applications including the NIST Privacy Engineering Objectives (PEOs) of predictability, manageability, and disassociability [9] as well as in higher level descriptions such as in the Fair Information Practice Principles (variations of which form a widely used basis for data privacy, such as the Organization for Economic Cooperation and Development (OECD) privacy guidelines [10]).

## 1.5 Risk Modeling

Risk modeling applies to both privacy and cybersecurity. Cybersecurity risk modeling centers on protecting organizations, whereas privacy focuses on individuals and groups. While realized privacy risks can include negative effects on an organization, their primary impacts are on people. Privacy risks are highly contextual because individuals and groups vary in their perceptions, preferences, and understanding of privacy and the complex systems that influence them.

Risk modeling identifies a range of potential risks for evaluation. A risk arises when a threat exploits a vulnerability, leading to an adverse outcome. However, not every threat will exploit every potential vulnerability. While each element of risk modeling—threats, vulnerabilities, and consequences—can be analyzed individually; threat modeling focuses specifically on understanding the threat component. To maximize the applicability of this paper's workflow (sequencing genomic material), the process focuses on threats instead of risks. In this paper, a genomic data threat related to privacy is any circumstance or event with the potential to compromise the predictability, manageability, and/or disassociability[7] of

[4] https://iapp.org/resources/glossary/#bodily-privacy

[5] https://iapp.org/resources/glossary/#information-privacy

[6] Note that concepts of privacy apply to people, not things. The term "data privacy" is not intended to imply that data has privacy; rather, the term refers to privacy as it relates to data processing and the impacts that data processing may have on people.

[7] These are the NIST privacy engineering objectives and are intended to be analogous to the fundamental cybersecurity properties of confidentiality, integrity, and availability. Predictability enables, "reliable assumptions

systems involving data associated with individuals (adapted from the NIST Privacy Framework [4] and NIST IR 8062 [9]). Note that genomic data privacy threats are distinct from the adverse consequences that could result from such compromises and can arise without external factors.

## 1.6 Threat Modeling

Threat modeling can support a broad stakeholder base who can then integrate the results into their larger and more specific risk modeling and management efforts.

The NCCoE team used the Four Question Framework, illustrated in the Appendix Figure 1, to structure the threat modeling process by answering:

1) "What are we working on?"
2) "What could go wrong?"
3) "What are we going to do about it?"
4) "Did we do a good job?"

In each of the four questions, "we" refers to the team performing the threat modeling. Though the questions are listed in sequential order, the process is iterative. Each question is addressed through specific techniques outlined in this paper. Answers to one question may be used to modify previous answers or highlight the incompleteness of an answer to a previous question. Threat modeling results improve through each iteration and should be conducted throughout the system's life cycle and whenever changes in the environment may impact threats. NIST CSWP 35 [7] demonstrates how the Four-Question Framework can be applied to cybersecurity threat modeling of a genomic data sequencing workflow.

Appendix C provides details for each tool used in this exercise with important details provided in this subsection. Threat modeling tools used in this exercise include the following:

1. NIST PRAM [1]: NIST's Privacy Engineering Program produced the Privacy Risk Assessment Methodology for identifying system privacy risks. Figure 3 shows the four PRAM worksheets: 1) Framing Business Objectives & Organizational Privacy Governance, 2) Assessing System Design (includes a separate Supporting Data Map), 3) Prioritizing Risk, and 4) Selecting Controls. The PRAM also leverages a non-exhaustive privacy risk model consisting of

by individuals, owners, and operators about data and their processing by a system." Manageability provides, "the capability for granular administration of data including alteration, deletion, and selective disclosure." Disassociability enables, "processing of data or events without association to individuals or devices beyond the operational requirements of the system."

"Problematic Data Actions" that may result in adverse effects for individuals listed in "Problems for Individuals."

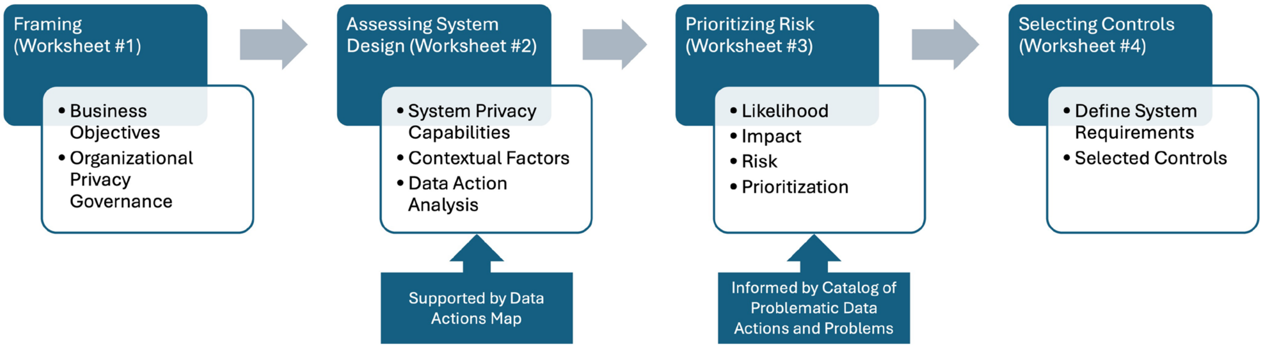

Framing (Worksheet #1)
Business Objectives
Organizational Privacy Governance
Assessing System Design (Worksheet #2)
System Privacy Capabilities
Contextual Factors
Data Action Analysis
Supported by Data Actions Map
Prioritizing Risk (Worksheet #3)
Likelihood
Impact
Risk
Prioritization
Informed by Catalog of Problematic Data Actions and Problems
Selecting Controls (Worksheet #4)
Define System Requirements
Selected Controls

**Figure 3. Overview of the NIST PRAM**

2. LINDDUN: A threat modeling tool for privacy inspired by the cybersecurity threat modeling tool STRIDE [5], the name is an acronym comprising seven different threat types: Linking, Identifying, Non-repudiation, Detecting, Data disclosure, Unawareness and Unintervenability, and Non-compliance. This technique relies on Dataflow Diagrams, which are useful for data privacy analysis and understanding the data life cycle.
3. PANOPTIC: A privacy analog to MITRE ATT&CK, the Pattern and Action Nomenclature of Privacy Threats in Context, was created based on real-world privacy attacks drawn from multiple sources. PANOPTIC has two closely related taxonomies of Contextual Domains and Privacy Activities that are enumerated in Table 23 and 24 of Appendix C.

# 2 Genomic Data Threat Modeling Example

## 2.1 Question 1: "What are we working on?"

Answering Question 1 helps teams identify activities and describe the system(s) being developed or analyzed. Because privacy is contextual, it is important to explicitly document that context in terms of the system and its surrounding environment. With this initial context, which may change over time, a more formalized description of system operation can be developed. The context is captured in a semi-structured fashion while augmented and annotated DFDs are used for the operational description.

The NCCoE Genomic Data Cybersecurity and Privacy project team documented the context and operational parameters by reviewing the workflow described in Figure 1, interviewing associated personnel, analyzing architecture documents, and building out the workflow to develop a shared understanding of the system environment, components, functionality, and interfaces. Through this process, the team established a baseline understanding to support analyzing genomic data threats regarding privacy and identifying potential interventions.

### 2.1.1 Context

For this analysis, context is considered along the broad dimensions of system and environmental. Relatedly, the NIST PRAM introduces the term contextual factors including system, individual, and organizational [1]. Systems typically exist in a larger environment of requirements or expectations. At the same time, systems will reflect environmental context with certain privacy commitments, approaches, and goals. An understanding of the environmental and system dimensions is necessary to provide a basis for threat modeling, especially for interpretations and judgments involved in determining what could go wrong (Section 2.2).

### 2.1.2 Environmental Context

NIST Privacy Risk Assessment Methodology (PRAM) [1] Worksheet 1 (Framing Business Objectives & Organizational Privacy Governance) is used together with elements of an adapted Worksheet 2 (Assessing System Design) to capture environmental context, primarily from the perspective of the sequencing service. Worksheet 1 focuses on the implementing organization(s) and is divided into two tasks: (1) frame organizational objectives and (2) frame organizational privacy governance, each of which consists of a series of questions and free form answers. Task 1 addresses business objectives and functional capabilities while Task 2 accounts for the governance structure that informs, enables, and constrains the system. These are environmental concerns because even though in principle they manifest themselves through the system, they are conditions that are external to the system.

Table 1 presents the Worksheet 1, Task 1 questions and responses. Table 2 presents the Task 2 questions and responses. Note that questions in Task 1 address overarching need and goals; responses therefore pertain to the *complete* example rather than solely the *core*.

**Table 1. PRAM Worksheet 1, Framing Business Objectives & Organizational Privacy Governance: Task 1 Questions and Responses**

| **1. Describe the mission/business needs that your system/product/service serves.** |
|---|
| Clinical Pipeline<br>Participating entities need to:<br>• Treat patients and provide genetic counseling<br>• Sequence their DNA to generate clinical results<br>• Deliver results to the patient and physician while ensuring patient privacy<br>Research Pipeline<br>Participating entities need to:<br>• Sequence provided DNA to generate research insights<br>• Deliver results to trusted research entity |
| **2. Describe the functional needs or capabilities of your system/product/service.** |
| Clinical Pipeline<br>Clinicians need:<br>• Sample intake protections and procedures (clinical form, test request form or TRF)<br>Sequencing service needs to:<br>• Maintain a proper chain of custody of the sample and associated data |

- Ensure the confidentiality of all patients by securing their data at rest using appropriate encryption
- Use proper bioinformatics data analysis pipelines that do not leak private data
- Ensure the privacy of patients by securing their in-transit data using appropriate encryption
- Securely disseminate results
- Retain or properly destroy data
- Maintain consent

Research Pipeline

Sequencing service needs to:

- Maintain consent to research
- Maintain a proper chain of custody of the sample and associated data
- Ensure the privacy of all direct data subjects by securing their data at rest
- Use proper bioinformatics tools that do not leak private data
- Ensure the privacy of direct data subjects by securing their in-transit data
- Securely disseminate results
- Retain or properly destroy data

**3. Describe any privacy-preserving goals for your system/product/service that you may plan to highlight or market to users or customers.**

Clinical Pipeline

Sequencing service will:

- Pseudonymize patient data while engaging in sequencing activities
- Preserve the privacy of patients and protect their data throughout the clinical pipeline

Research Pipeline

Sequencing service will:

- Pseudonymize direct data subjects' data to relevant standards (e.g., HIPAA Safe Harbor or expert determination)
- Preserve the privacy of direct data subjects and protect their data throughout the research pipeline
- Protect research results (e.g., treatment personalization approach) within the research pipeline

Table 2. PRAM Worksheet 1, Framing Business Objectives & Organizational Privacy Governance: Task 2 Questions and Responses

| **1. Legal Environment: Identify any privacy-related statutory, regulatory, contractual and/or other frameworks within which the organization must operate. List any specific privacy requirements.** |
|---|
| Include:<br>• Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule, including Protected Health Information (PHI)[8]<br>• Genetic Information Nondiscrimination Act of 2008<br>• Clinical Laboratory Improvement Amendments (CLIA)[9]<br>• College of American Pathologists (CAP)[10]<br>• European Union General Data Protection Regulation (GDPR)<br>• State (e.g., California Consumer Privacy Act, Alabama HB21 Genetic Data)<br>• Applicable National Institutes of Health (NIH) requirements and regulations<br>• The Common Rule (45 CFR 46, U.S.) – Federal regulations that:<br>○ Mandate Institutional Review Board (IRB) oversight<br>○ Require informed consent procedures<br>○ Provide additional protections for vulnerable groups like children and prisoners<br>• Grant-specific privacy requirements |
| **2. Identify any privacy-related principles or other commitments to which the organization adheres (e.g., Fair Information Practice Principles, Privacy by Design principles, ethics principles).** |
| • Accreditation requirements (CLIA/CAP)<br>• NIH Data User Code of Conduct[11]<br>• Food and Drug Administration (FDA) Genomic Sampling and Management of Genomic Data Guidance for Industry[12]<br>• Medical and research ethics (IRB)<br>• Good clinical practice (GCP) |
| **3. Identify any privacy goals that are explicit or implicit in the organization's vision and/or mission.** |
| • Ensure the privacy of all individuals by protecting their data |
| **4. Identify any privacy-related policies or statements within the organization, or business unit.** |
| • Limit sharing of individuals' data by limiting access to only those with a need to know |

[8] Protected information is defined by the HIPAA Privacy rule as all "individually identifiable health information." https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

[9] https://www.cms.gov/medicare/quality/clinical-laboratory-improvement-amendments

[10] https://www.cap.org/laboratory-improvement/accreditation/laboratory-accreditation-program

[11] https://sharing.nih.gov/accessing-data/accessing-genomic-data/using-genomic-data-responsibly/genomic-data-user-code-of-conduct#for-users-accessing-data-on-or-after-january-25,-2025

[12] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e18-genomic-sampling-and-management-genomic-data-guidance-industry

| |
|---|
| • Vet privacy practices of third parties who are used for outside services and hosting<br>• Keep all privacy training documents up to date as well as ensure staff regularly receive training<br>• Handling policies of samples and data reflect privacy obligations |
| **5. Document your organization’s risk tolerance with respect to privacy from your organization’s enterprise risk management strategy.** |
| The following are considered untenable:<br>• Risk from third parties absent specific legal constraints<br>• Individuals’ data are mixed with data or entered into systems not directly related to sample processing (e.g., administrative) |

PRAM Worksheet 2 (Assessing System Design) captures contextual factors that go beyond the organization itself, situating it within the larger environment and in relation to affected individuals. Table 3 presents the organizational contextual factors for the clinical and research use cases while Table 4 presents the contextual factors for individuals. As with Worksheet 1, these apply to the complete example.

**Table 3. PRAM Worksheet 2, Assessing System Design: Organizational Contextual Factors**

| Clinical Use Case |
|---|
| Organizations include a private clinic or other healthcare provider and a non-profit genomic sequencing/bioinformatics laboratory in this example |
| Public perception: Especially high expectation of privacy for all organizations handling genomic data in a clinical setting |
| Relationships: Patient has no pre-existing relationship with the genomic sequencing/bioinformatics laboratory and has interacted with the private clinic or other healthcare provider by providing their data/sample along with their consent for use of the data/sample for clinical testing |
| **Research Use Case** |
| Organizations include a national research organization and a non-profit genomic sequencing/bioinformatics laboratory in this example |
| Public perception: High expectation of privacy for all organizations handling genomic data |
| Relationships: Data subject has no pre-existing relationship with the genomic sequencing/bioinformatics laboratory and has interacted with the national research organization by providing their data/sample along with their consent for use of the data/sample for research |

**Table 4. Worksheet 2, Assessing System Design: Contextual Factors for Individuals**

| Clinical Use Case |
|---|
| High sensitivity about genomic data/physical samples; individual and their relatives could all be affected |
| Patients’ levels of technical sophistication and understanding of genomic sequencing and how it is used in clinical care decisions vary widely |

| |
|---|
| Potential patient misunderstanding regarding what organization(s) will have access to their genomic data when providing additional consent for research |
| Potential patient misunderstanding regarding personal and familial impacts of genomic data |
| **Research Use Case** |
| High sensitivity about genomic data/physical samples; individual and their relatives could all be affected |
| Data subjects’ levels of technical sophistication and understanding of genomic research vary widely |
| Potential direct data subject misunderstanding regarding what organization(s) will have access to their genomic data when providing initial consent for research |
| Pseudonymized or acceptable aggregate research results are intended to be made public, according to the specifics of the consent provided by direct data subjects |

### 2.1.3 System Context

The team described the system context using two complementary approaches: an adapted PRAM Worksheet 2 and the PANOPTIC Contextual Domains. PRAM Worksheet 2 addresses system privacy capabilities and other contextual factors for the complete example. As a controlled taxonomy, PANOPTIC provides a structured and granular description of system context for the *complete* example, including categories of data, that complements the information captured by Worksheet 2. Worksheet 2 addresses system privacy capabilities and other contextual factors for the complete example. System capabilities—in terms of the PEOs of predictability, manageability, and disassociability—are presented in Table 5 and Table 6 for the clinical and research use cases respectively. Worksheet 2 contextual factors are presented in Table7 for both the clinical and research use cases.

**Table 5. PRAM Worksheet 2, Assessing System Design: System Privacy Capabilities for Clinical Use Case**

| Predictability |
|---|
| Patient’s data is only used for clinical efforts according to the specifics of their provided consent |
| Patient’s data is appropriately pseudonymized during sequencing service use |
| **Manageability** |
| Patient is able to provide consent for their data to be used that specifies the type(s) of clinical uses that are consented to |
| Patient can, at any time, request information about how their data is being used for clinical purposes |
| Patient can, at any time, withdraw consent for their data being used for clinical purposes |
| **Disassociability** |
| Digital genomic data provided for clinical uses have been pseudonymized, allowing for the data to be used in the lab without associating the genomic data directly with a patient |

Table 6. PRAM Worksheet 2, Assessing System Design: System Privacy Capabilities for Research Use Case

| Predictability |
|---|
| Direct data subject’s data is only used for research efforts according to the specifics of their provided consent |
| Direct data subject’s data is pseudonymized prior to use in research or acceptable aggregate statistics are used in research |
| **Manageability** |
| Direct data subject is able to provide consent for their data to be used for research, including more fine-grained consent, if desired, that specifies the type(s) of research that are consented to |
| Direct data subject can, at any time, request information about how their data is being used for research |
| Direct data subject can, at any time, withdraw consent for their data being used for research |
| **Disassociability** |
| Digital genomic data provided for research has had direct identifiers removed and cannot be analyzed at the individual subject level, allowing for the data to be used for research projects without associating the genomic data with the direct data subject |
| Research results do not include genomic data that could be analyzed at the individual subject level |
| The non-profit sequencing service can carry out research tasks and analyses without associating a direct data subject with the provided sample |
| The national research organization can review the results provided by the non-profit sequencing service and will not be able to connect them back to a direct data subject |
| While the nature of genomic data makes complete disassociability impossible to guarantee, accepted practices – releasing results that cannot be analyzed at the individual subject level and maintaining direct subject data in controlled access repositories - are used to allow research use of genomic data |
| Digital genomic data provided for research have been pseudonymized and cannot be analyzed at the individual subject level, allowing for the data to be used for research projects without associating the genomic data with a direct data subject |

Table 7. PRAM Worksheet 2, Assessing System Design: System Contextual Factors

| Clinical Use Case |
|---|
| System includes a private clinic or other healthcare provider and a non-profit genomic sequencing/bioinformatics laboratory |
| Privacy policies governs system |
| Public perception: Especially high expectation of privacy for all organizations handling genomic data in a clinical setting |

| |
|---|
| Relationships: Patient has no pre-existing relationship with the genomic sequencing/bioinformatics laboratory and has interacted with the private clinic or other healthcare provider by providing their data/sample along with their consent for use of the data/sample for clinical testing |
| **Research Use Case** |
| Research results not containing identifiable information are intended to be made public, according to the specifics of the consent provided by the direct data subjects |
| History with system: Direct data subject has already provided to the national research organization their data/sample along with consent for research use of the data/sample; data subject has no pre-existing relationship with the sequencing service; system has similarity to other publicly funded genomics research systems |
| Two parties involved: One public, one non-profit |
| Genomic sequencing/bioinformatics lab may use third party bioinformatics tools during data analysis if required to produce the necessary data for the research project |

Similarly, separate PANOPTIC contextual mappings were constructed for the clinical and research use cases. We present these textually in Table 8 and Table 9 rather than in their original graphical forms, which can be found in Appendix D.

**Table 8. PANOPTIC Contextual Mapping for Clinical Use Case**

| Contextual Domain | Contextual Element/ Sub-element | PANOPTIC Definition | Comment |
|---|---|---|---|
| Environment | PC01.01 Digital | Data action in a digital environment | |
| Environment | PC01.02 Physical | Data action in a physical environment, including physical processes such as filling out a paper form | |
| Distribution | PC02.02 One to one | Data custodian shares information with one other entity | |
| Distribution | PC02.03 One to many | Data custodian shares information with a discrete number of other entities[13] | |
| Interaction | PC03.01.01 No interaction | Data subject does not directly interact at all with the entity or their proxy | Applies to indirect data subjects |

[13] Note that this entry and the rest in this column of corresponding tables is a definition from PANOPTIC used to identify scope and context for analysis

| Contextual Domain | Contextual Element/ Sub-element | PANOPTIC Definition | Comment |
|---|---|---|---|
| Interaction | PC03.02.02 Discrete proxy interaction | Data subject's proxy interacts a discrete number of times, including once, with the entity or their proxy | Genetic sample is considered a data proxy for the direct data subject |
| Engagement | PC04.01.08 Genetics | Data subjects who, based on the differentiating characteristic of genetics, are within a contextually sensitive population | Pertains to specific genetic traits, such as susceptibility to particular diseases or other health conditions |
| Engagement | PC04.01.10 Illness or injury | Data subjects who, based on the differentiating characteristic of their health status, are within a contextually sensitive population | |
| Engagement | PC04.01.11 Other context-specific populations | Data subjects who, based on the differentiating characteristic of another context-specific population, are within a contextually sensitive population | Relates to population-specific diseases or health conditions |
| Data Type | PC05.02 Demographic | Population characteristics of the data subject, e.g., education level, ethnicity, religion, citizenship | Some of these data may be part of the patient's health record |
| Data Type | PC05.06 Contact information | Information including the identity of, and the means to communicate with, the associated data subject(s) | |
| Data Type | PC.05.07 Health | Information pertaining to the data subject's health status, including mental health, or use of health-related products or services | |
| Data Type | PC05.08 Financial | Information pertaining to the data subject's financial status or transactions, e.g., credit ratings and history, income, bank accounts | These data pertain to billing and insurance |
| Data Type | PC05.15.01 Persistent direct identifier | A consistent identifier that one can be reasonably confident directly associates data with the data subject, such as a name | |
| Data Type | PC05.15.02 Persistent pseudo-identifier | An identifier that enables data to be repeatedly associated with the same data subject(s) or their proxy without knowing their identity, such as a username or a MAC address | Pertains to sample pseudonymization during sequencing service processing |

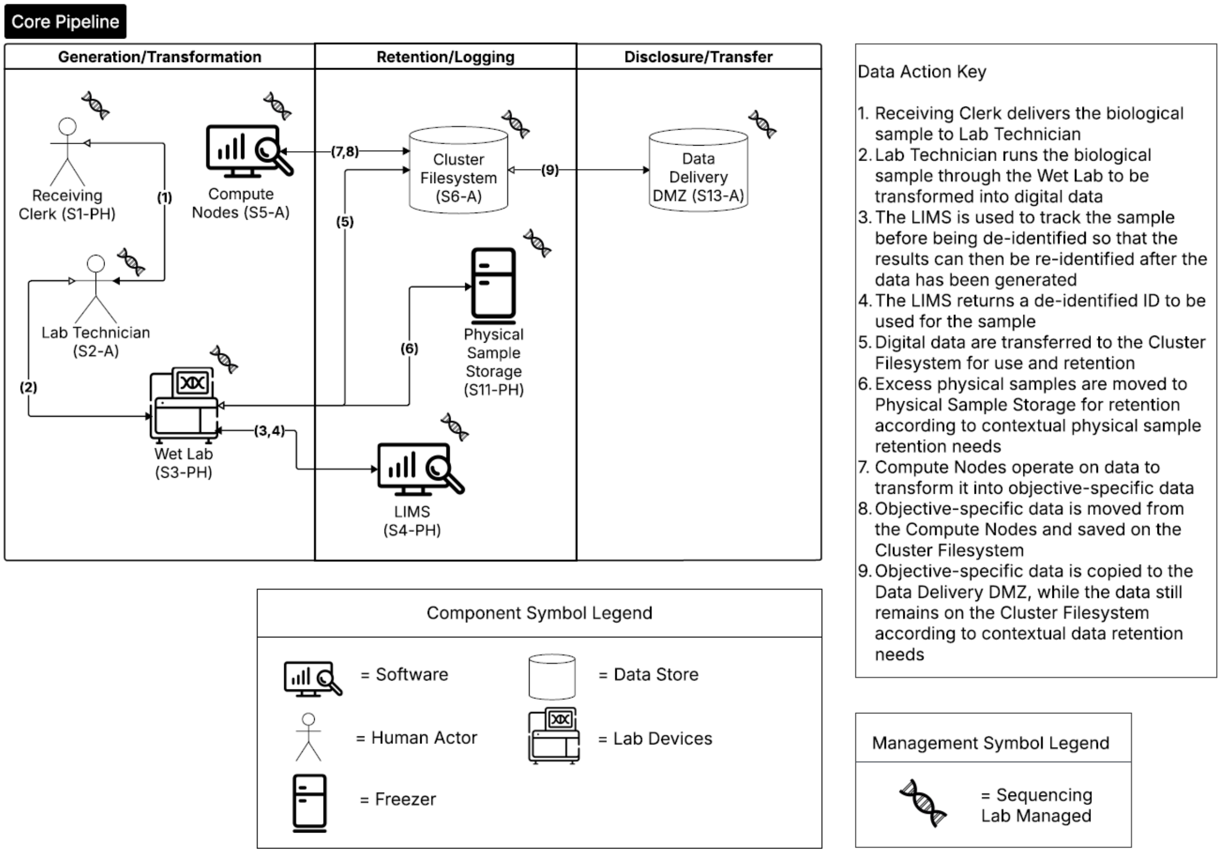
429 **Table 9. PANOPTIC Contextual Mapping for Research Use Case**

| Contextual Domain | Contextual Element/ Sub-element | PANOPTIC Definition | Comment |
|---|---|---|---|
| Environment | PC01.01 Digital | Data action in a digital environment | |
| Environment | PC01.02 Physical | Data action in a physical environment, including physical processes such as filling out a paper form | |
| Distribution | PC02.03 One to many | Data custodian shares information with a discrete number of other entities | Approved project collaborators analyzing data |
| Interaction | PC03.01.01 No interaction | Data subject does not directly interact at all with the entity or their proxy | Applies to indirect data subjects |
| Interaction | PC03.02.02 Discrete proxy interaction | Data subject's proxy interacts a discrete number of times, including once, with the entity or their proxy | Genetic sample is considered a data proxy for the direct data subject |
| Engagement | PC04.01.01 Age | Data subjects who, based on the differentiating characteristic of age, are within a contextually sensitive population | Relates to the focus of some research studies, if explicit in recruitment and/or analysis plan |
| Engagement | PC04.01.02 Race & ethnicity | Data subjects who, based on the differentiating characteristic of race and/or ethnicity, are within a contextually sensitive population | Relates to the focus of some research studies, if explicit in recruitment and/or analysis plan |
| Engagement | PC04.01.05 Gender | Data subjects who, based on the differentiating characteristic of gender, are within a contextually sensitive population | Relates to the focus of some research studies, if explicit in recruitment and/or analysis plan |
| Engagement | PC04.01.08 Genetics | Data subjects who, based on the differentiating characteristic of genetics, are within a contextually sensitive population | Pertains to specific genetic traits, such as susceptibility to particular diseases or other health conditions |
| Engagement | PC04.01.10 Illness or injury | Data subjects who, based on the differentiating characteristic of their health status, are within a contextually sensitive population | |
| Engagement | PC04.01.11 Other context-specific populations | Data subjects who, based on the differentiating characteristic of another context-specific | Relates to the focus of some research studies |

| Contextual Domain | Contextual Element/ Sub-element | PANOPTIC Definition | Comment |
|---|---|---|---|
| | | population, are within a contextually sensitive population | |
| Data Type | PC05.02 Demographic | Population characteristics of the data subject, e.g., education level, ethnicity, religion, citizenship | |
| Data Type | PC.05.07 Health | Information pertaining to the data subject's health status, including mental health, or use of health-related products or services | |
| Data Type | PC05.13.01 Preferences | Information pertaining to the data subject's interests or favor of one alternative over another | Pertains to options regarding particular types of research |
| Data Type | PC05.15.01 Persistent direct identifier | A consistent identifier that one can be reasonably confident directly associates data with the data subject, such as a name | |
| Data Type | PC05.15.02 Persistent pseudo-identifier | An identifier that enables data to be repeatedly associated with the same data subject(s) or their proxy without knowing their identity, such as a username or a MAC address | Pertains to sample pseudonymization |

### 2.1.4 Operational Description

This section describes system operations and data using augmented and annotated dataflow diagrams as described in Appendix E, https://pages.nist.gov/nccoe-genomic-data-threat-modeling/Vol_C/Appendix/appendixE.html#dataflow-diagram-legend. Figure 4 shows the DFD for the *core* example: common elements of the clinical and research use cases in a generalized version of their shared dataflows. This is followed by descriptions of the diagraming techniques and the diagram itself. Complete diagrams, including the dataflow diagram symbol legend, covering the clinical and research use cases can be found in Appendix E. Note that in the research, use case digitized rather than physical samples may be shared with the sequencing service.



Data Action Key

1. Receiving Clerk delivers the biological sample to Lab Technician
2. Lab Technician runs the biological sample through the Wet Lab to be transformed into digital data
3. The LIMS is used to track the sample before being de-identified so that the results can then be re-identified after the data has been generated
4. The LIMS returns a de-identified ID to be used for the sample
5. Digital data are transferred to the Cluster Filesystem for use and retention
6. Excess physical samples are moved to Physical Sample Storage for retention according to contextual physical sample retention needs
7. Compute Nodes operate on data to transform it into objective-specific data
8. Objective-specific data is moved from the Compute Nodes and saved on the Cluster Filesystem
9. Objective-specific data is copied to the Data Delivery DMZ, while the data still remains on the Cluster Filesystem according to contextual data retention needs

**Figure 4. Core Example Dataflow Diagram**

DFDs depict communication paths among components of the system being analyzed, which provide information important to any analysis of data privacy. DFDs also help teams produce a common architecture document that can be used for other collaboration and development activities outside the threat modeling effort.

To address privacy, this notation was altered and augmented in several ways. First, components were assigned more informative symbols as well as unique identifiers. All symbols are identified in the Component Symbol Legend of the diagrams. The identifiers include a prefix and a suffix, with the prefix indicating which use case the component belongs to. Because the *core* example DFD is, by definition, a shared dataflow, the “S” prefix is used in all cases. (In the full analysis, “C” and “R” are used to indicate the clinical and research use cases respectively. Also, because these are drawn from the complete example, the numbering is not fully sequential.) The suffix indicates more specific sub-case(s), including potentially all, in which the component participates. Delineating these is optional but can aid interpretation by further contextualizing components based on their roles.

Second, each component was annotated with a management symbol indicating the responsible party. These are identified in the Management Symbol Legend of the diagrams. Note, that in the *core* example DFD a single party, the sequencing service, is responsible for all elements. Third, each dataflow was numbered, and its purpose described in the Data Action Key. Bidirectional dataflows were assigned two numbers to account for the dataflow in each direction.

The last type of modification, though, is the most critical for privacy. That alteration bears on how the elements of the DFDs are organized. The elements are arranged to fall into columns that relate to different types of data actions. Data actions describe what is happening to data and reflect different stages of the information life cycle. These can vary somewhat in their particulars and the data actions employed here are those discussed in the NIST PRAM [1] and the NIST Privacy Framework [4]. Table 10 lists these along with their descriptions.

**Table 10. Data Action Types and Dataflow examples**

| Action Types | Dataflow Examples |
|---|---|
| Collection | Data are ingested by a component. |
| Generation/Transformation | Data are processed to produce further data or to clean/manipulate/unify the data. |
| Disclosure/Transfer | Data are revealed or communicated to others. This action is disclosure when the data moves from one managing entity to another and transfer when it moves between components managed by the same entity. |
| Retention/Logging | Data and/or metadata are stored for future use. |
| Disposal | Data are destroyed or otherwise rendered inaccessible. |

The *core* example DFD includes three types of data actions: Generation/Transformation, Disclosure/Transfer, and Retention/Logging. To begin the pipeline, the Receiving Clerk obtains the sample to be sequenced and provides it to the Lab Technician who will prepare and transform it into digital data with the systems present within the Wet Lab. During this process, the laboratory information management system (LIMS) catalogs the sample and provides a pseudo-identifier for future tracking. The leftover sample material is properly stored within the Physical Sample Storage while the digital data are moved from the sequencer to the Cluster Filesystem. The data on the Cluster Filesystem are sent to the Compute Nodes for analysis before the returned information is sent back to the Cluster Filesystem and ultimately uploaded to the Data Delivery demilitarized zone (DMZ). These dataflows and actions are present for all use cases in which a genomic sequencing service may carry out sequencing projects.

## 2.2 Question 2: “What could go wrong?”

At this point environmental and system context have been captured and the operational dataflows and actions have been documented. The analytical processes of genomic data threat modeling for privacy must now be applied to these descriptions. Those processes consist of two principal activities: (1) dataflow analysis to identify threats and (2) threat alignment and validation.

To address “what could go wrong,” the dataflow analysis (1) employed LINDDUN and its catalog of threat trees to associate potential genomic data threats regarding privacy with specific dataflows and actions, then (2) created PANOPTIC attack mappings for the genomic sequencing workflow. Both models are needed because the LINDDUN analysis identifies abstract threats that are theoretically possible while PANOPTIC identifies steps that could form a practical attack. Where practical attack and theoretical threat align, the combination is validated against the NIST PEOs. This exercise ensures that potential threats are both conceivable and executable, and that these would impact at least one of the NIST PEOs.

### 2.2.1 LINDDUN Analysis

The LINDDUN [2] methodology involves assessing each distinct dataflow for potential threats. A dataflow consists of a source, the flow itself, and a destination. To avoid confusion, we refer to this triad as a dataflow segment. Using the modified Assess System Design table in PRAM Worksheet 2, each dataflow segment in the *core* example DFD (Figure 4) is documented. In addition to the source, flow, and destination, the applicable data actions[14] are also noted. Each dataflow segment is also assigned a purpose (based on the Data Action Key) using the Context column.

For each documented dataflow segment, relevant LINDDUN threats are then identified, using as a starting point the mapping of segment-based high-level threat types, shown in Table 11. This mapping is a heuristic for determining potential LINDDUN threats and involved components. The LINDDUN threat trees [2] that detail those threat types can then be used to determine whether and which more granular threats potentially apply to that segment based on its constituent elements and context. Those threats judged potentially applicable are captured in the LINDDUN Analysis column in Table 12, including the scenario. Note that multiple threats may apply to a single dataflow segment.

**Table 11. LINDDUN Per Element Threat Mapping Heuristic**

| Source (Src) | Destination (Dst) | L | I | NR | D | DD | U | NC |
|---|---|---|---|---|---|---|---|---|
| Process | Process | Src-flow-Dst | Src-flow-Dst | Src-flow-Dst | Src-flow | Src-flow-Dst | Src-Dst | Src-Dst |
| Process | Store | Src-flow-Dst | Src-flow-Dst | Src-flow-Dst | Src-flow | Src-flow-Dst | Src-Dst | Src-Dst |
| Process | External | Src-flow-Dst | Src-flow-Dst | Src-flow-Dst | Src-flow | Src-flow-Dst | Src-Dst | Src-Dst |
| Store | Process | Src-flow-Dst | Src-flow-Dst | Src-flow-Dst | Src-flow | Src-flow-Dst | Src-Dst | Src-Dst |
| External | Process | Src-flow-Dst | Src-flow-Dst | Src-flow-Dst | Src-flow | Src-flow-Dst | Src-Dst | Dst |

To illustrate, consider dataflow segment Number 1 in Table 12. It consists of a receiving clerk delivering a physical biological sample to a lab technician for genomic sequencing. Leveraging the data action column helps us infer that this is a process-to-process segment. Consulting Table 11 and the threat

[14] While the diagram organizes the nodes by data action, dataflow segments may involve more than a single data action.

definitions, as well as the context of the segment, we conclude that linking is the only relevant threat. Sending samples to a technician known to be associated with work on a particular disease could link the samples to that disease, an instance of L.2.2.1, profiling an individual. The other possibilities can be dismissed because at this stage:

- The sample must still be associated with the direct data subject as part of the workflow
- There is nothing for the direct data subject to repudiate, aside from providing the sample to those who must necessarily be aware that the sample has been provided
- Because the sample must be identifiable, detection is unavoidable
- The only data disclosure is inherent in the workflow and therefore unproblematic
- The direct data subject has provided informed consent and is aware of their options
- Standard practices are being employed in the workflow

The process proceeds similarly for the remaining eight dataflow segments, resulting in Table 12. Note that a segment can be subject to more than one threat, as is the case for segment 8.

Table 12. LINDDUN Dataflow Analysis for the Core Exam

| No. | Source | Dataflow Type | Data Action 1 | Data Action 2 | Destination | Context (purpose) | LINDDUN Analysis (applicable threats) | |
|---|---|---|---|---|---|---|---|---|
| 1 | Receiving Clerk (S1-PH) | Physical Sample | Transfer | | Lab Tech (S2-A) | Send physical sample to lab tech for research project | L.2.2.1 | Sending samples to wet lab known to be researching a specific disease at that time could link samples to that disease |
| 2 | Lab Tech (S2-A) | Physical Sample | Transfer | | Wet Lab (S3-PH) | Send physical sample to wet lab for sequencing | L.2.2.1 | Sending samples to wet lab known to be researching a specific disease at that time could link samples to that disease |
| 3 | Wet Lab (S3-PH) | Physical Sample | Transfer | Retention | Physical Sample Storage (S11-PH) | Send physical sample for storage in appropriate freezers | L.2.1.2 | Sending group of X samples together to freezers around the same time as a project known to be doing Y disease research could link the samples to Y disease |
| 4 | Wet Lab (S3-PH) | Sample Metadata | Generation | Retention | LIMS (S4-PH) | Generate pseudonymized ID to be used for sample | I.2.1.1 | Nature of genomic data makes complete disassociability impossible to guarantee |
| 5 | LIMS (S4-PH) | Sample Metadata | Transfer | | Wet Lab (S3-PH) | Send back to wet lab the pseudonymized ID to be used for sample | L.2.1.2 | Samples put into LIMS around same time could receive IDs with linkable characteristics, which then allows linkage of sample group to a study around same time, unless LIMS is cautious of this |

| No. | Source | Dataflow Type | Data Action 1 | Data Action 2 | Destination | Context (purpose) | LINDDUN Analysis (applicable threats) | |
|---|---|---|---|---|---|---|---|---|
| 6 | Wet Lab (S3-PH) | Sequence Data | Transfer | Retention | Cluster Filesystem (S6-A) | Send digital sequence data to be stored | L.2.1.2 | Samples that are put into the cluster filesystem around the same time could be interpreted as being linked to a study about Y disease around the same time |
| 7 | Cluster Filesystem (S6-A) | Sequence Data | Transfer | | Compute Nodes (S5-A) | Send digital sequence data to Compute Nodes to operate on digital sequence data to transform it into objective-specific data | L.2.1.2 | Samples sent to compute nodes around same time could be interpreted as being linked to a study about Y disease around same time |
| 8 | Compute Nodes (S5-A) | Sequence Data, Context-relevant Research Data | Transformation | | Cluster Filesystem (S6-A) | Operate on sequence data to create context-relevant research data | DD.4.1.2 | Bioinformatics tools come from a variety of developers that can change over time; corruption within this supply chain, especially if left unmonitored, could result in research subject data being disclosed |
| | | | | | | | U.1.1 | Data subject does not clearly understand what data actions that analysis tools along the pipeline will perform on their data |

| No. | Source | Dataflow Type | Data Action 1 | Data Action 2 | Destination | Context (purpose) | LINDDUN Analysis (applicable threats) | |
|---|---|---|---|---|---|---|---|---|
| 9 | Compute Nodes (S5-A)<br>Cluster Filesystem (S6-A) | Sequence Data, Context-relevant Research Data<br>Context-relevant Research Data | Transformation<br>Transfer | | Cluster Filesystem (S6-A)<br>Data Delivery DMZ (S13-A) | Operate on sequence data to create context-relevant research data<br>Send generated context-relevant research data to data delivery DMZ for to make it available for delivery | L.2.1.2 | Samples that are put into the data delivery DMZ around the same time could be interpreted as being linked to a study about Y disease around the same time |

The complete LINDDUN analysis can be found in Appendix E. Note that for manageability the analysis was initially divided into clinical, research, and shared use cases, the last based on the common portion of the two use cases. The results were then combined into a single system design table. This table was then sorted on the specific LINDDUN threats.

### 2.2.2 PANOPTIC Analysis

The LINDDUN analysis identifies potential threats at the level of dataflows. However, real-world privacy attacks are not typically launched at that level, nor do they consist of a single self-contained element. They are less abstract and operate at the system level. The PANOPTIC analysis is a necessary complement to the LINDDUN analysis as it will describe potential threats from a system perspective. The LINDDUN analysis is then used to determine whether the threats identified at the dataflow level support the projected attacks as described by PANOPTIC. If not, the PANOPTIC attacks are considered non-actionable.

While the LINDDUN analysis is grounded in system specifics as captured by DFDs, the PANOPTIC analysis involves actively imagining in practical terms what might take place. Utilizing the PANOPTIC Privacy Activities mapping template, a privacy attack mapping for the *core* example was generated. Table 13 lists the threat actions identified for the *core* example based on high-level knowledge of the system and its context. The complete PANOPTIC mappings for the clinical and research use cases are provided in Appendix E.

540 **Table 13. Threat Actions Identified by the PANOPTIC Privacy Activity Mapping for the Core Example**

| PANOPTIC Threat Action | Definition | Elaboration |
|---|---|---|
| PA02.02 Consent: Imprecise | Key data actions are not presented clearly enough to constitute informed consent | May not provide details on how research is conducted, and which parts of the pipeline are privacy-relevant |
| PA03.09 Collection: Recording | Capturing a physical or digital artifact representing an aspect or likeness of the data subject | |
| PA03.11 Collection: Biological sample | Collecting biological materials or specimens (e.g., blood, urine, tissue cells, or saliva) from the data subject | |
| PA05.01.01 Identification: Re-identification | Re-associating data with the data subject that had been treated to remove those associations | |
| PA05.02.02 Identification: Pseudo-identifier | Assigning a pseudo-identifier (e.g., randomly generated ID) | |
| PA07.01 Manageability: No individual access to information | The data subject or their proxy cannot obtain or view their collected personal data | |
| PA07.02 Manageability: No individual management of information content | The data subject or their proxy cannot transform (e.g., move, copy, edit) their collected personal data | Direct data subject cannot change their data that is used for research |
| PA07.03 Manageability: No individual deletion of information | The data subject or their proxy cannot delete their collected personal data | Once the research data is published, the direct data subject cannot remove theirs from the body of research |
| PA07.05 No individual control of information use | The data subject or their proxy cannot control how their information is used | Direct data subject cannot manage what types of research studies use their data |
| PA08.01.01 Aggregation: Single source profiling | Assembling and organizing data points about specific data subjects from a single source | The research project must determine whether or not a given direct data subject exhibits the trait being studied, implying profiling with the single source being their provided sample |
| PA08.02.01 Aggregation: Single source clustering | Assembling and organizing data points regarding groups of people from a single source | Research studies may look for commonalities across genomic samples |
| PA08.02.02 Aggregation: Multi-source clustering | Assembling and organizing data points regarding groups of people from multiple sources | Research studies may seek insights on a specific population potentially characterized along |

| PANOPTIC Threat Action | Definition | Elaboration |
|---|---|---|
| | | multiple dimensions, implying clustering |
| PA09.01.01 Processing: Deriving information about individuals | Determining or extracting novel information about the data subject by analyzing information | Research project must determine if the trait being studied is exhibited by the data subject |
| PA09.01.02 Processing: Deriving aggregate information | Determining or extracting novel aggregate information by analyzing information | Research project may seek insights about a given population regarding a genetic trait |
| PA09.01.03 Processing: Deriving sensitive information | Determining or extracting novel sensitive information by analyzing information | Genetic information and insights gained can be sensitive information |
| PA09.01.04 Processing: Deriving derogatory information | Determining or extracting novel derogatory information by analyzing information | Genetic diseases or susceptibility to them can be considered derogatory information |
| PA09.03 Processing: Introducing bias | Data action is adversely influenced by bias | Bias could be introduced into research projects if the demographic spread of the data pool is not balanced. (This may not be possible for some studies, such as one targeting a trait only present in a specific population.) |
| PA10.01 Sharing: Affording revelations | Making available information that enables the discovery of further information | A research project that a direct data subject joins may yield results now or in the future, including the relevance of the research topic for the data subject |
| PA11.01 Use: Implication | Establishing a particularized derogatory suspicion or accusation regarding the data subject | |
| PA12.01 Retention & destruction: Data not destroyed after use | Information has not been disposed at the conclusion of its life cycle | May be indeterminate for research data |
| PA12.02 Retention & destruction: Data improperly destroyed | Information remains at least partially recoverable despite attempts to destroy it | Flow cell insufficiently cleaned and sequencer supply chain not cleaning hard drives |

Table 14 describes five attack scenarios that are specific to the *core* example. Each scenario was determined by considering how specific threat actions could be used by an actor as part of an attack involving a distinct DFD segment. Since attacks could apply to different DFD segments, the table in some cases associates multiple identical attacks with the same scenario. Appendix F provides the comprehensive analysis that was performed on the complete example, which includes all the Attack Numbers and Scenario IDs. Table 14 extracts only the attack scenarios relevant to the *core* example, aligning with the Attack Numbers, Scenario IDs, and Privacy Threat Actions from the comprehensive analysis found in Appendix F.

Table 14. Attack Scenarios Relevant to the Core Example

| Attack Numbers from Complete Example | Scenario ID | PANOPTIC Threat Actions Describing the Attack | Scenario Description |
|---|---|---|---|
| 1, 14, 15 | S1.1 | PA03.09, PA03.11, PA08.01.01, PA10.01, PA11.01 | Pipeline actor uses physical access to correlate study details with physical samples and associated metadata. |
| 2-5 | S1.2 | PA03.09, PA05.02.02, PA08.02.02, PA10.01, PA11.01 | Pipeline actor uses physical access to correlate study details with digital data. |
| 26 | S6 | PA05.01.01 | Pipeline actor uses digital access to correlate study details with digital data. |
| 55 | S6 | PA03.09, PA09.01.01, PA09.01.03, PA09.01.04, PA11.01 | Pipeline actor uses digital access to correlate study details with digital data. |
| 65 | S17 | PA02.02, PA07.05 | Sequencing service staff utilizes third party tools and software that may perform additional data actions unbeknownst to a direct data subject.[15] |

In the first scenario described in Table 14, attack numbers 1, 14, and 15, which constitute health status inference attacks, can be broken down as follows: The attack involves an actor with a role in the sequencing pipeline physically accessing artifacts relating to direct data subjects (PA03.09, Collection: Recording) in the form of biological samples (PA03.11) and their associated metadata (as per PC05). The actor can correlate the research studies that will use these samples with the samples and their metadata (PA08.01.01, Aggregation: Profiling: Single source profiling), which may reveal other information, such as potential susceptibility to a particular disease (PA10.01, Sharing: Affording revelations). This would enable the attacker to discern something negative about the individual's health status (PA11.01, Use: Implication).

[15] Further discussion of this issue can be found in the NIST Quick-Start Guides for Cybersecurity Supply Chain Risk Management (https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1305.pdf) and Due Diligence Assessment (https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1326.ipd.pdf).

### 2.2.3 Threat Validation

As previously indicated, threat validation consists of two steps: mapping PANOPTIC attacks to relevant LINDDUN threats and mapping LINDDUN-validated attacks against the NIST PEOs of predictability, manageability, and disassociability. If a PANOPTIC attack does not align with one or more LINDDUN threats or if an aligned attack does not appear to undermine at least one of the PEOs, then the threat is invalid and removed from further consideration during this modeling process iteration.

Validation of PANOPTIC attacks against LINDDUN threats amounts to assessing the relationship between the threat actions that constitute the attack and the relevant LINDDUN threats. In most cases, that relationship is many-to-many. Therefore, carrying out this assessment involves judgement informed by the surrounding context. To facilitate this determination, Appendix G includes a mapping between PANOPTIC threat actions and LINDDUN threats in both directions. Because such mappings exist in all cases, the mere existence of a potentially relevant LINDDUN threat is insufficient validation.

For attacks aligned with LINDDUN threats, validation against the PEOs serves to confirm that the attacks actually met the definition of a threat put forward in Section 1.4 by potentially undermining system predictability, manageability, and/or disassociability. Some attacks may impact more than one PEO, but a validated attack must impact at least one.

Table 15 lists the validation results for the five attack scenarios relevant to the *core* example from Table 14. These were extracted from the complete combined validation table found in Appendix G. This table documents the LINDDUN Analysis and PEOs impacted by the threat, aligned to the Attack Number, the Scenario ID, PANOPTIC Threat Action, and LINDDUN Threat.

**Table 15. Core Example Attack Validations**

| Attack Number | Scenario ID | PANOPTIC Threat Action | LINDDUN Threat | LINDDUN Analysis | Impacted PEOs |
|---|---|---|---|---|---|
| 1 | S1.1 | PA03.09, PA03.11, PA08.02.01, PA10.01, PA11.01 | L.2.1.2 | Sending the group of X samples together to the freezers around the same time as a project known to be doing Y disease research could link the samples to Y disease | Predictability |
| 2 | S1.2 | PA03.09, PA05.02.02, PA08.02.02, PA10.01, PA11.01 | L.2.1.2 | Samples that are put into the LIMS around the same time could receive IDs with linkable characteristics, which then allows linkage of the sample group to a study around the same time, unless the LIMS implements measures to prevent this | Predictability |
| 3 | S1.2 | PA03.09, PA05.02.02, PA08.02.02, | L.2.1.2 | Samples that are put into the cluster filesystem around the same time could be interpreted as being linked to a | Predictability |

| Attack Number | Scenario ID | PANOPTIC Threat Action | LINDDUN Threat | LINDDUN Analysis | Impacted PEOs |
|---|---|---|---|---|---|
| | | PA10.01,<br>PA11.01 | | study about Y disease around the same time | |
| 4 | S1.2 | PA03.09,<br>PA05.02.02,<br>PA08.02.02,<br>PA10.01,<br>PA11.01 | L.2.1.2 | Samples sent to the compute nodes around the same time could be interpreted as being linked to a study about Y disease around the same time | Predictability |
| 5 | S1.2 | PA03.09,<br>PA05.02.02,<br>PA08.02.02,<br>PA10.01,<br>PA11.01 | L.2.1.2 | Samples that are put into the data delivery DMZ around the same time could be interpreted as being linked to a study about Y disease around the same time | Predictability |
| 14 | S1.1 | PA03.09,<br>PA03.11,<br>PA08.01.01,<br>PA10.01,<br>PA11.01 | L.2.2.1 | Sending samples to the technician known to be researching a specific disease could link the samples to that disease | Predictability<br>Disassociability |
| 15 | S1.1 | PA03.09,<br>PA03.11,<br>PA08.01.01,<br>PA10.01,<br>PA11.01 | L.2.2.1 | Sending samples to the wet lab known to be researching a specific disease at that time could link the samples to that disease | Predictability<br>Disassociability |
| 26 | S6 | PA05.01.01 | I.2.1.1 | Nature of genomic data makes complete disassociability impossible to guarantee | Predictability<br>Disassociability |
| 55 | S6 | PA03.09,<br>PA09.01.01,<br>PA09.01.03,<br>PA09.01.04,<br>PA11.01 | DD.4.1.2 | Bioinformatics tools come from a variety of developers that can change over time; corruption within this supply chain, especially if left unmonitored, could result in research subject data being disclosed | Predictability |
| 65 | S17 | PA02.02,<br>PA07.05 | U.1.1 | Data subject does not clearly understand what data actions that analysis tools along the pipeline will perform on their data | Predictability<br>Manageability |

To understand the validation process, consider attack number 14 as a specific example from Table 15. The PANOPTIC threat actions and sub-actions that make up the attack map to the LINDDUN threat types of Linking, Non-repudiation, Detecting, and Data Disclosure. (Definitions of these are provided in Appendix C.) Neither Non-repudiation nor Detecting is relevant to this scenario and can be dropped from consideration. By sorting the dataflow analysis table (Table 12) on the LINDDUN threat designators it is then possible to review the dataflows related to Linking and Data Disclosure. Matching scenario components are then identified by sorting on the Dataflow column to group those entries involving physical samples.[16] The dataflow analysis for the *core* example contains multiple instances involving physical samples susceptible to threat L2.2.1, profiling an individual. This validates attack 14 against the LINDDUN analysis. Based on both the LINDDUN threat and the PANOPTIC threat actions (profiling and revelation in particular), attack 14 clearly undermines predictability as well as disassociability, validating it against the PEOs. Therefore, we can conclude that this is a valid threat.

As Table 15 indicates, all PANOPTIC attacks were successfully validated against LINDDUN threats and the LINDDUN-supported attacks validated against the PEOs. As a result, all the threats are candidates for responses.

## 2.3 Question 3: "What are we going to do about it?"

Once threats have been validated, decisions must be made regarding how to respond. The high-level options for addressing validated threats align with the options for risk management:

1. **Eliminate**. This is the most desired outcome; however, it is often challenging and may involve forgoing a specific feature or functionality. For example, in the case of attack number 1 in the *core* example (Table 15), removing the receiving clerk from the pipeline by sending physical samples directly to the relevant lab technician would introduce logistical complications that could prove infeasible. If a feature or function is required to accomplish one of the use case's Mission Objectives (MOs), then eliminating the threat is not possible.
2. **Disrupt**. This involves identifying, adding, and/or improving controls to frustrate attacks. For example, the nexus of attack number 1 in the *core* example is single source profiling. Controls targeting this threat action would disrupt the entire attack. This is explored in more detail in Section 3.
3. **Transfer Responsibility**. This strategy transfers responsibility for addressing the threat to another entity, who may have resources of their own to intervene or who can better tolerate the presence of the threat. Documentation of this responsibility transfer and appropriate agreements are an important aspect for implementing this option.
4. **Accept**. In any system, there are threats which are challenging or impossible to disrupt but whose presence is judged to be tolerable. For example, attack number 55 in the *core* example reflects potential issues of software supply chains. However, the system design may be judged sufficiently robust to warrant accepting the threat of using externally developed software, which may then be paired with threat intelligence monitoring. These accepted threats need to be documented and periodically reviewed, tolerance for accepting threats may change over time.

[16] In the *core* example the number of dataflows and associated threats is so limited that no sorting is necessary. In contrast, the complete example contains almost 100 itemized LINDDUN threats.

When working on Question 3, it is important to consider all four options: eliminate, disrupt, transfer, accept. The impact on the mission posed by the threat, as well as the organization's threat (and risk) tolerance, will guide decision-making. The most common and perhaps most complex response is to disrupt the threat by applying additional controls or reconfiguring existing ones. There may be multiple interventions (potentially ranging across eliminate, disrupt, and transfer) for a threat with varying costs and effectiveness. Choices should be guided by the organization's mission, tolerances, and resources.

If interventions (i.e., responses other than accept), are chosen, they need to be adequately documented to be implementable. There should be sufficient detail to support implementation and testing. If a threat is accepted, the reasoning and assumptions should be documented. In all cases, decisions should be periodically revisited as both the environment and the organization's risk tolerance may change over time.

This section describes the process of determining threat responses, using the *core* example as an illustration. In selecting and implementing interventions it is important to consider responsibility, verifiability (preferably automated), maintainability, and usability. All systems will inevitably need updates and modifications; the managing party for verification and maintenance of each intervention needs to be clearly defined.

### 2.3.1 Threat Prioritization

While a small number of threats can be easily prioritized by inspection, typical analyses require a more systematic approach. Therefore, each attack was characterized in terms of its feasibility and difficulty. In this exercise the *core* example only exhibits a handful of validated threats though the *complete* example identifies close to 100 as described in Appendix E, G and F with many falling below a prioritization threshold.

Assigning values to attack feasibility (how credible it is) and difficulty (how hard it is to execute) is inherently subjective. Feasibility reflects a number of factors, including threat actor opportunity, capability, and (in the case of people and organizations) motivation. It is assessed using one of three designations: plausible, implausible, or indeterminate. In those cases where the system itself is the threat actor and the attack is intrinsic to normal system operation, of course, the attack is not only feasible but pre-determined and therefore plausible by definition. Attack difficulty reflects the capabilities and resources required, as well as how many discrete threat actions are involved, and is indicated using a five-step scale. The difficulty scale and context-specific criteria based on the state of the data are given in Table 16.

**Table 16. Attack Difficulty Scale**

| Difficulty Level | Context-specific Data Criteria |
|---|---|
| Negligible | Analyzed results with additional input, professional recommendations, patient PHI |
| Minor | Tool output, clinical reports without additional analysis |
| Moderate | Physical samples or raw sequencing data |
| Significant | Metadata produced by the sequencing service devices or sample requests that are pseudonymized |
| Severe | Information about the machines and analysis tools used or the staff that works there |

Returning to attack number 14, this is a plausible attack based on the necessity of the dataflow through an honest but potentially curious threat actor, the receiving clerk. The attack is of moderate difficulty since it revolves around physical samples and requires knowledge and correlation of certain information. Table 17 shows the characterization for the entire *core* example. Table 13 lists descriptions of individual PANOPTIC threat actions.

**Table 17. Core Example Threat Characteristics**

| No. | PANOPTIC Attack | LINDDUN Threat | LINDDUN Analysis | Feasibility | Difficulty |
|---|---|---|---|---|---|
| 1 | PA03.09, PA03.11, PA08.02.01, PA10.01, PA11.01 | L.2.1.2 | Sending the group of X samples together to the freezers around the same time as a project known to be doing Y disease research could link the samples to Y disease | Plausible | Moderate |
| 2 | PA03.09, PA05.02.02, PA08.02.02, PA10.01, PA11.01 | L.2.1.2 | Samples that are put into the LIMS around the same time could receive IDs with linkable characteristics, which then allows linkage of the sample group to a study around the same time, unless the LIMS is cautious of this | Plausible | Significant |
| 3 | PA03.09, PA05.02.02, PA08.02.02, PA10.01, PA11.01 | L.2.1.2 | Samples that are put into the cluster filesystem around the same time could be interpreted as being linked to a study about Y disease around the same time | Plausible | Moderate |
| 4 | PA03.09, PA05.02.02, PA08.02.02, PA10.01, PA11.01 | L.2.1.2 | Samples sent to the compute nodes around the same time could be interpreted as being linked to a study about Y disease around the same time | Plausible | Moderate |
| 5 | PA03.09, PA05.02.02, PA08.02.02, PA10.01, PA11.01 | L.2.1.2 | Samples that are put into the data delivery DMZ around the same time could be interpreted as being linked to a study about Y disease around the same time | Plausible | Minor |
| 14 | PA03.09, PA03.11, PA08.01.01, PA10.01, PA11.01 | L.2.2.1 | Sending samples to the technician known to be researching a specific disease could link the samples to that disease | Plausible | Moderate |
| 15 | PA03.09, PA03.11, PA08.01.01, PA10.01, PA11.01 | L.2.2.1 | Sending samples to the wet lab known to be researching a specific disease at that time could link the samples to that disease | Plausible | Moderate |

| No. | PANOPTIC Attack | LINDDUN Threat | LINDDUN Analysis | Feasibility | Difficulty |
|---|---|---|---|---|---|
| 26 | PA05.01.01 | I.2.1.1 | Nature of genomic data makes complete disassociability impossible to guarantee | Plausible | Moderate |
| 55 | PA03.09,<br>PA09.01.01,<br>PA09.01.03,<br>PA09.01.04,<br>PA11.01 | DD.4.1.2 | Bioinformatics tools come from a variety of developers that can change over time; corruption within this supply chain, especially if left unmonitored, could result in research subject data being disclosed | Plausible | Minor |
| 65 | PA02.02,<br>PA07.05 | U.1.1 | Data subject does not clearly understand what data actions that analysis tools along the pipeline will perform on their data | Plausible | Minor |

Once all validated threats have had feasibility and difficulty values assigned, the different combinations can be assigned normalized numerical values for ranking purposes, as shown in Table 18. Plausible attacks of negligible difficulty carry the highest value (resulting in higher priority) while implausible attacks of severe difficulty carry the lowest value (resulting in lower priority). To incorporate additional nuance into the rankings, weights were assigned to the LINDDUN threat types to reflect their relative severity in the context of genomic sequencing, as shown in Table 19. Note, though, that these values are purely an ordering mechanism and do not have any intrinsic meaning.

**Table 18. Attack Feasibility and Difficulty Combination Values**

| Difficulty<br>Feasibility | Negligible | Minor | Moderate | Significant | Severe |
|---|---|---|---|---|---|
| Plausible | 1.0 | 0.8 | 0.6 | 0.4 | 0.2 |
| Indeterminate | 0.9 | 0.7 | 0.5 | 0.3 | 0.1 |
| Implausible | 0.8 | 0.6 | 0.4 | 0.2 | 0.0 |

**Table 19. LINDDUN Threat Weights**

| LINDDUN Threat Type | Weight |
|---|---|
| Data Disclosure | 1.0 |
| Identifying | 0.85 |
| Linking | 0.7 |
| Non-compliance | 0.5 |
| Unawareness and Unintervenability | 0.5 |
| Detecting | 0.3 |
| Non-repudiation | 0.2 |

These values and weights were multiplied for each attack and the results used to rank order the threats in the *core* example from highest to lowest priority, as shown in Table 20. (Ties are resolved using attack number.) The prioritization of threats for the complete example is provided in Appendix G.

**Table 20. Core Example Threats in Ranked Order from Highest to Lowest Priority**

| No. | LINDDUN Threat | Feasibility | Difficulty | Ranking Value |
|---|---|---|---|---|
| 55 | DD.4.1.2 | Plausible | Minor | 0.80 |
| 5 | L.2.1.2 | Plausible | Minor | 0.56 |
| 26 | I.2.1.1 | Plausible | Moderate | 0.51 |
| 1 | L.2.1.2 | Plausible | Moderate | 0.42 |
| 3 | L.2.1.2 | Plausible | Moderate | 0.42 |
| 4 | L2.1.2 | Plausible | Moderate | 0.42 |
| 14 | L.2.2.1 | Plausible | Moderate | 0.42 |
| 15 | L.2.2.1 | Plausible | Moderate | 0.42 |
| 65 | U.1.1 | Plausible | Minor | 0.40 |
| 2 | L.2.1.2 | Plausible | Significant | 0.28 |

Given the limited number of threats in the *core* example, it would be reasonable to explicitly consider a response to each threat, including the option of acceptance. However, given that the number of threats in the complete example is an order of magnitude larger, some organizations may opt to accept threats below a certain priority threshold without further deliberation. Determining that threshold is a function of organizational tolerances and resources.

### 2.3.2 Response Determination

High-priority threats tend to readily give rise to decisions to intervene (typically in the form of elimination or disruption). Likewise, low-priority threats tend to prompt decisions to accept the threat. In contrast, determining the appropriate response to threats occupying the middle ground—such as attack number 14—is often less straightforward.

Attack number 14 involves a seemingly unavoidable dataflow, so simply eliminating the dataflow is not an option, nor is there any obvious way of transferring responsibility. This leaves the option of either accepting the presence of the threat or disrupting it. Determining which course to pursue may require first exploring disruption options so that their viability may be considered.

There are several reference sources for such controls, but one of the most prominent is NIST Special Publication (SP) 800-53r5, Security and Privacy Controls for Information Systems and Organizations [6]. However, different organizations may have varying resources and expertise for selecting controls and control enhancements relevant to given threats. Though organizations may have different approaches to this process, the following describes a way of facilitating the process to map from individual PANOPTIC threat actions to candidate controls using the NIST Privacy Framework, leveraging NIST's crosswalk[17] from PF Subcategories to 800-53 controls.

Handling a large number of candidate controls, even after duplicates are accounted for, requires a reduction step. One way of further constraining the effort is to focus on critical PANOPTIC threat actions. These are threat actions that others are dependent upon; disrupting critical threat actions in effect invalidates the attack. In attack number 14, the critical threat action is single source profiling. The threat actions that enable it (Recording and Biological sample) are unavoidable while the remaining threat actions (Affording revelations and Implication) are enabled by it. Focusing on single source profiling (and its associated LINDDUN threat) results in a set of less than 20 candidate controls. Appendix C shows this winnowing process, starting from the two PF Categories implicated by this threat action, mapping from the Categories to the relevant Subcategories, and from the Subcategories to the relevant 800-53 controls.

Each Subcategory is augmented with an ordered tuple (e.g., [1 2 1 1]), representing the priority of that Subcategory for each of the four selected MOs drawn from the *Genomic Data Profile* [5] (Organizational Tailoring in Appendix C provides more details of this approach). These tuples can be used to prioritize potential controls that might be employed to disrupt threats given that the Genomic Data Profile provides a list of MOs for organizations processing genomic data and prioritizes PF Subcategories (or outcomes) to support achieving those MOs. Based on the genomic sequencing workflow, four relevant MOs were selected:

MO 2: Manage privacy risk to existing and future relatives

[17] https://github.com/usnistgov/PrivacyFrmwkResources/raw/master/resources/NIST%20SP%20800-53%20Crosswalk/csf-pf-to-sp800-53r5-mappings.xlsx

MO 3: Identify, model, and address cybersecurity and privacy risks of processing genomic data

MO 5: Manage privacy risk to donors

MO 12: Promote the use of privacy-enhancing technologies as well as secure technologies for sharing genomic data

Each Privacy Framework Subcategory includes this tuple that indicates the Genomic Data Profile prioritization of MO 2, MO 3, MO 5, and MO 12 listed as [1 2 1 2].

**Table 21. Mapping from Single Source Profiling to SP 800-53r5 Controls**

| Privacy Framework Function - Category | Privacy Framework Subcategory | 800-53 Controls | 800-53 Control Family |
|---|---|---|---|
| Control-P – Disassociated Processing | CT.DP-P2: Data are processed to limit the identification of individuals [1 2 1 2] | AC-23 | Access Control |
| Control-P – Disassociated Processing | CT.DP-P2: Data are processed to limit the identification of individuals [1 2 1 2] | AU-3(3) | Audit and Accountability |
| Control-P – Disassociated Processing | CT.DP-P2: Data are processed to limit the identification of individuals [1 2 1 2] | IA-4(8) | Identification and Authentication |
| Control-P – Disassociated Processing | CT.DP-P2: Data are processed to limit the identification of individuals [1 2 1 2] | PE-8(3) | Physical and Environmental Protection |
| Control-P – Disassociated Processing | CT.DP-P2: Data are processed to limit the identification of individuals [1 2 1 2] | SA-8(33) | System and Services Acquisition |
| Control-P – Disassociated Processing | CT.DP-P2: Data are processed to limit the identification of individuals [1 2 1 2] | SI-12(1)<br>SI-12(2)<br>SI-19 | System and Information Integrity |
| Control-P – Disassociated Processing | CT.DP-P3: Data are processed to limit the formulation of inferences about individuals’ behavior or activities [2 3 2 2] | AC-23 | Access Control |
| Control-P – Disassociated Processing | CT.DP-P3: Data are processed to limit the formulation of inferences about individuals’ behavior or activities [2 3 2 2] | AU-16(3) | Audit and Accountability |
| Control-P – Disassociated Processing | CT.DP-P3: Data are processed to limit the formulation of inferences about individuals’ behavior or activities [2 3 2 2] | IA-8(6) | Identification and Authentication |
| Control-P – Disassociated Processing | CT.DP-P3: Data are processed to limit the formulation of inferences about individuals’ behavior or activities [2 3 2 2] | PL-8 | Planning |

| Privacy Framework Function - Category | Privacy Framework Subcategory | 800-53 Controls | 800-53 Control Family |
|---|---|---|---|
| Control-P – Disassociated Processing | CT.DP-P3: Data are processed to limit the formulation of inferences about individuals' behavior or activities [2 3 2 2] | PM-7 | Program Management |
| Control-P – Disassociated Processing | CT.DP-P3: Data are processed to limit the formulation of inferences about individuals' behavior or activities [2 3 2 2] | SA-8(33)<br>SA-17 | System and Services Acquisition |
| Control-P – Disassociated Processing | CT.DP-P3: Data are processed to limit the formulation of inferences about individuals' behavior or activities [2 3 2 2] | SC-2(2) | System and Communications Protection |
| Control-P – Disassociated Processing | CT.DP-P3: Data are processed to limit the formulation of inferences about individuals' behavior or activities [2 3 2 2] | SI-19 | System and Information Integrity |
| Protect-P – Protective Technology | PR.PT-P2: The principle of least functionality is incorporated by configuring systems to provide only essential capabilities [3 2 2 2] | AC-3 | Access Control |
| Protect-P – Protective Technology | PR.PT-P2: The principle of least functionality is incorporated by configuring systems to provide only essential capabilities [3 2 2 2] | CM-7 | Configuration Management |

Once the set of potentially applicable controls has been narrowed down in this way, the tuples derived from MO 2, MO 3, MO 5, and MO 12 can be used to prioritize the Subcategories and by extension control selection.[18] MO 2, which deals with privacy risk to relatives, is not relevant for this attack and can be ignored. MO 12, which addresses use of privacy-enhancing technologies (PETs), assigns the same priority to all three Subcategories and can also be ignored as it does not contribute any differentiation. The prioritizations for MO 3 and MO 5, however, readily yield an ordering of (1) CT.D-P2 [2 1], (2) PR.PT-P2 [2 2], (3) CT.DP-P3 [3 2].

Reviewing the controls associated with CT.DP-P2 for those that appear most relevant or impactful, we find two candidates:

- IA-4(8) Pairwise Pseudonymous Identifiers – Generate pairwise pseudonymous identifiers.

[18] While in principle the Mission Objectives could be employed to prioritize threats rather than controls, the MOs selected for this workflow provide insufficient differentiation; MOs 3, 5, and 12 will be implicated by most threats.

- SI-12(1) Limit Personally Identifiable Information Elements − Limit personally identifiable information being processed in the information life cycle to the following elements of PII: [Assignment: organization-defined elements of personally identifiable information].

Reviewing the controls associated with PR.PT-P2, we find:

- CM-7 Least Functionality − Configure the system to provide only [Assignment: organization-defined mission essential capabilities].

Finally, reviewing the controls associated with CT.DP-P3, we find:

- AU-16(3) Disassociability − Implement [Assignment: organization-defined measures] to disassociate individuals from audit information transmitted across organizational boundaries.

All of these in various ways could help prevent the association of identifiable individuals with specific studies or tests. However, the most direct ones are arguably those associated with the highest priority Subcategory, CT.DP-P2. Both of these point toward the need to break the link between specific individuals and (inferred) specific lab operations. Employing pairwise pseudonymous identifiers as per IA-4(8) (generating a unique identifier for every sample, even if the samples pertain to the same data subject) could accomplish this if samples could be pseudonymized at the source. This would involve the client interacting with a sequencing service system (possibly via an application programming interface, API) to generate the pseudonymous identifier that would be used for shipping purposes. The receiving clerk would then enter/scan the identifier into the system, but with restricted access to information (SI-12(1)) and functionality (CM-7), to determine which lab technician should receive it. This approach could possibly leverage an existing interface (e.g., a Web portal used for communicating results).

While in this case one might well have arrived at the same or similar conclusions without the Subcategory prioritization, some of the threat actions map to a significantly greater number of Subcategories with a much larger set of associated controls. In those cases, Subcategory prioritization can provide beneficial structure that facilitates control selection. Where there are many Subcategories, prioritization might even provide a basis for limiting control selection to those associated with the higher-priority Subcategories.

It stands to reason that similar threats should respond to similar interventions, so in principle these disruptions should be applicable to all instances of the scenario, addressing attacks 1 through 5 as well as attack number 15. This also applies to other intervention types. One might also potentially identify similar attacks in different scenarios by searching on the associated critical threat actions and/or specific LINDDUN threats. Further, selection of controls that show up frequently across disruptions may offer greater cost-effectiveness, as long as care is taken to ensure that all targeted threats are sufficiently addressed. Also, given that some threats involve 3rd parties, controls that focus on agreements or those such as CA-02 Control Assessments and CA-03 Information Exchange may offer interventions that address multiple threats.

## 2.4 Question 4: "Did we do a good job?"

Question 4, "Did we do a good job?" directs the project team to evaluate the effectiveness of answers to Questions 1-3. This paper outlines the effort to document genomic data processing environments from a privacy standpoint (Question 1), identify genomic data threats to privacy (Question 2), and implement interventions (Question 3). The threat modeling process is designed to be iterative. This paper showcases the process rather than an exhaustive analysis to guide other teams conducting genomic data threat modeling for privacy on their own systems. Question 4 also helps emphasize that this process will be repeated to address changes in the system and threat environments.

This section provides guidelines on how to address Question 4 and suggests additional activities that can be used by teams to evaluate their efforts. The threat modeling documentation in the form of adapted PRAM Worksheets 1 and 2 should be reviewed and updated periodically to address new threats, system changes, new assumptions, and changes in risk tolerance.

### 2.4.1 Did We Do a Good Job Documenting the System and Its Data Actions?

Section 2.1 documents the system context, including PANOPTIC Contextual Domain mappings and DFDs. DFDs directly support threat identification and analysis while Contextual Domain mappings indirectly support the process. In comparison to cybersecurity threat modeling, threats related to privacy can arise from systems operating as designed therefore trust boundaries are a concept that is not used. As such, the entirety of the system is potentially relevant for privacy analysis.

The following activities could potentially improve the documentation of the system and its data actions:

- Review the system scope to ensure that it has captured the full breadth of data and data actions.
- Check whether contextual information is sufficiently specific.
- Check whether DFDs are sufficiently detailed to capture communications between systems and all data actions.
- Review documentation and information from suppliers, developers, and users—including that addressing data subject consents and preferences—to consider any updates required.
- Review change control processes to ensure that changes are documented properly.
- Update the documentation to reflect changes to the system context or dataflows, including system interconnections, devices added, or issues identified through testing or monitoring.
- Review data handling processes to ensure adherence to best practices and reflect any changes in the contextual information or DFDs as applicable.

### 2.4.2 Did We Do a Good Job Identifying and Documenting Threats?

To answer, “Did we do a good job?” on Question 2, “What could go wrong?” the project team evaluated whether the threat model adequately identified and documented threats to data subjects. Section 2.2 enumerates the threats identified for the *core* example based on the LINDDUN dataflow analysis and the PANOPTIC attacks, with the results for the complete example documented in Appendix F. The following actions could improve threat identification.

**Evaluate the comprehensiveness of the LINDDUN analysis**. The LINDDUN per element threat mapping heuristic shown in Table 11 acts as a completeness check. With this table, a completeness check can be done for the typical threats against external entities, processes, data stores, and dataflows. If there are possible threats that were not considered by the team, this highlights an area for additional consideration. Maintaining a checklist for each dataflow segment by making and marking copies of Table 11 could help prevent potentially relevant threats from being overlooked.

**Evaluate the comprehensiveness of the identified PANOPTIC threat actions**. When evaluating the PANOPTIC attacks:

- Consider privacy attacks that have occurred in the genomic stakeholder community and closely adjacent industries. Threat intelligence can be used to identify attacks favored by actors who are known to target an industry. The Bioeconomy Information Sharing and Analysis Center (BIO-ISAC) is one potential source of such intelligence.[19]
- Consider whether the identified scenarios and selected PANOPTIC threat actions reflect these attacks, or if additional scenarios and/or threat actions should be considered.
- Determine whether the threats being considered adequately reflect the threats listed in published documents for the genomic community, such as NIST IR 8432 [11].

**Review and confirm that invalidated threats are in fact invalid.** Revisit invalidated threats to ensure that they were not mistakenly invalidated because a relevant LINDDUN threat was overlooked, or a relevant PEO was not recognized as such. To facilitate such a review, it is essential to retain documentation of invalidated threats.

**Review organizational policies, strategies, and processes to determine if there are other threat areas not being addressed by the technical evaluation.** Such a review may uncover otherwise overlooked but relevant activities or scenarios. For example, sharing of data for ancillary purposes could take place via mechanisms largely separate from the target system, such as copying of LIMS logs. If the sequencing service is actively seeking to be acquired, that could potentially present threats related to any retained samples or data.

### 2.4.3 Did We Do a Good Job Responding to the Threats?

Section 2.3.2 discussed the kinds of responses to the identified threats that might be considered. More specifically, using one of the attacks in the *core* example, it illustrated how to intervene by disrupting the threat using standard controls by reasoning from the attack to particular controls by way of the NIST PF.

The following actions could evaluate and improve on this approach:

- Review interventions to assess how well they address the LINDDUN threats associated with the attacks.
- Expand interventions to cover additional PF Subcategories beyond those that were addressed based on the prioritizations in the Genomic Data Profile for the Mission Objectives that have been established. (See Organizational Tailoring in Appendix C.)
- Review the documentation from Question 1 to check which, if any, interventions may already be present.
- If the answers to Questions 1 and/or 2 have changed, revisit the relevant response determinations.

[19] https://www.isac.bio/

- Develop a surveillance plan that incorporates any findings from assessments, tabletop exercises, or ongoing vulnerability monitoring using available resources[20] and documents how they will be integrated into future threat modeling activities.

### 2.4.4 Additional Activities

The following additional actions help evaluate the thoroughness of responses and regularly consider the impact of any changes to the system or threat environment. A legal review may be appropriate to determine if the interventions, accepted threats, and transferred responsibilities (particularly the manner of transfer notification) meet the necessary regulatory requirements (GV.PO-P).

**Review Threat Responses.** Appropriate documentation of threat responses beyond disruption is critical to ensuring that they can be revisited as circumstances change.

- **Eliminate.** Eliminating threats often removes features. Accompanying documentation should justify the trade-offs involved. This documentation is necessary because threat models will need to be revisited as the system and organization evolves. Future threat modeling efforts may involve different participants who may not be familiar with the system and will rely on this documentation.
- **Accept.** Threats that are accepted should be documented sufficiently to explain why. For example, the interventions necessary to disrupt attack number 1 and similar attacks in the *core* example, where sending a group of samples together to the same technician could link the samples to the disease they're known to be researching, may be considered too onerous relative to the threat's priority. The reason for the threat acceptance needs to be documented so that if the process surrounding sample intake changes, the threat and the response to it can be reassessed.
- **Transfer.** When responsibility for threats is transferred, documentation should clearly indicate the entity assuming accountability for those threats. That entity may then choose to intervene, accept, or further transfer responsibility for the threat. Documentation adequately specifies the obligations and expectations of both parties. Note that not all responsibilities can be transferred, such as those which are legally obligated (e.g., breach notification).

**Update DFDs.** As interventions are added, DFDs may need to be updated. The possibility of new threats against changed or added elements should be considered. If new threats arise from changes, appropriate responses must be determined, which could include reconsidering the intervention.

**Review PANOPTIC Attacks.** If there are interventions in place that disrupt multiple common threat actions, that can be a positive indication of the layering of controls, which supports robust privacy protection.

[20] https://nvd.nist.gov

**Utilize Framework Profiles.** Teams can use the Genomic Data Profile to identify further interventions by considering additional priority Subcategories for each relevant Mission Objective. (See Organizational Tailoring in Appendix C.) Alternatively, PF Subcategories associated with the disruptions selected during Question 3 activities can be used to inform an organization's PF Target Profile, which could leverage a Community Profile such as the Genomic Data Profile. The organization can then identify potential gaps by comparing its Current Profile to its Target Profile.

**Track Interventions Throughout the System Life Cycle.** Threat interventions should be documented, reviewed, tested, and maintained as the threat environment changes. This may include the following considerations:

- During the implementation phase, threat modeling should be periodically revisited and updated. Consider whether the intervention caused problems and if so, what were the impacts.
- Once interventions are operational, consider their effectiveness and any unanticipated negative impact to Mission Objectives. For example, if the intervention reduced the ability of direct data subjects to exercise control over their data (CT.PO-P3), consider if the protection provided by the intervention justified that diminution of control.
- Organizations should update their threat response and possibly the relevant aspects of their threat model after an intervention fails, considering whether the failure resulted from erroneous analysis. Performing and documenting root cause analysis can usefully inform future decisions.
- Privacy assessment, including automated and manual red teaming, is another useful tool to evaluate how the interventions and threat modeling perform and how they can be improved.
- Tabletop and functional exercises as described in SP 800-84 [12] can also be very helpful in evaluating Question 3 performance and can be done both before and after a system is in use.

# 3 Conclusion

The paper provides an example of how a threat modeling process can be employed in a systematic and consistent manner to analyze genomic data threats related to privacy to the Clinical Client, Research Partner, and Genomic Sequencing Service environments. It shows how the process charts, characterizes, and analyzes the dataflows of each use case to identify specific types of potential threats, while describing possible actualizing attacks. It also demonstrates how valid threats can be prioritized and provides an illustrative example of how to identify and select threat-disrupting interventions.

This threat modeling process identified notable genomic data threats and concerns in the use cases examined. One key finding is the limited ability for individuals to exercise informed consent and maintain control over their genomic data as it moves across increasingly complex dataflows. Additionally, the interconnected nature of genomic data introduces the potential for direct subjects' data to impact indirect subjects, such as relatives, further complicating privacy management.

Additional details regarding our threat modeling approach, methodology, dataflows, mappings, and threat validation can be found in Appendices C-G. The scope of our analysis was constrained to two use cases and focused on dataflows between two organizations. Further analysis could explore the complexities of environments involving multiple entities and more intricate dataflows. Also, the rapidly evolving field of genomics, coupled with dynamic threat landscape, present considerations that could also be analyzed. Expanding the scope could yield additional insights into privacy challenges for genomic data processing. Organizations may also consider approaches for implementing ongoing threat monitoring to supplement threat modeling.

# Appendix A List of Acronyms

The following acronyms are used in this publication.

| | |
|---|---|
| **API** | Application Programming Interface |
| **ATT&CK** | Adversarial Tactics, Techniques & Common Knowledge |
| **BIO-ISAC** | Bioeconomy Information Sharing and Analysis Center |
| **CAP** | College of American Pathologists |
| **CLIA** | Clinical Laboratory Improvement Amendments |
| **DFD** | Dataflow Diagram |
| **DMZ** | Demilitarized Zone |
| **DNA** | Deoxyribonucleic acid |
| **FDA** | Food and Drug Administration |
| **GCP** | Good Clinical Practice |
| **GDPR** | EU General Data Protection Regulation |
| **GINA** | Genetic Information Nondiscrimination Act of 2008 |
| **HIPAA** | Health Insurance Portability and Accountability Act |
| **IR** | Internal Report |
| **IRB** | Institutional Review Board |
| **LIMS** | Laboratory Information Management System |
| **LINDDUN** | Linking, Identifying, Detecting, Data Disclosure, Unawareness and Unintervenability, and Non-compliance privacy threat types |
| **MO** | Mission Objective |
| **NCCoE** | National Cybersecurity Center of Excellence |
| **NIH** | National Institutes of Health |
| **NIST** | National Institute of Standards and Technology |
| **OSS** | Open-Source Software |
| **PANOPTIC** | Pattern and Action Nomenclature of Privacy Threats in Context |
| **PEO** | Privacy Engineering Objective |
| **PET** | Privacy-Enhancing Technology |
| **PF** | NIST Privacy Framework |
| **PRAM** | Privacy Risk Assessment Methodology |

| | |
|---|---|
| **SP** | NIST Special Publication |
| **STRIDE** | Spoofing, Tampering, Repudiation, Information Disclosure, and Elevation of Privilege cybersecurity threat types |
| **SQL** | Structured Query Language |
| **TRF** | Test Request Form |

# Appendix B References

[1] National Institute of Standards and Technology (2021) NIST Privacy Risk Assessment Methodology (PRAM). https://www.nist.gov/privacy-framework/nist-pram

[2] LINDDUN Privacy Threat Modeling. Available at https://linddun.org/

[3] MITRE PANOPTIC Privacy Threat Model. Available at https://ptmworkshop.gitlab.io/#/panoptic

[4] NIST Privacy Framework. https://www.nist.gov/privacy-framework

[5] Martin N, et al. (2023) Cybersecurity Framework Profile for Genomic Data. (National Institute of Standards and Technology, Gaithersburg, MD), Initial Public Draft NIST Interagency or Internal Report (IR) 8467. https://doi.org/10.6028/NIST.IR.8467.2pd

[6] National Institute of Standards and Technology (2025) Special Publication SP 800-53 Rev.5 https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final

[7] Pulivarti R, Wagner J, Zook, J, Kreider B, Wilson K, Snyder J, Wojtyniak M, Ross S, Whitlow P, Sheldon J, Brown I, Pape P, Alim E (2024) Cybersecurity Threat Modeling the Genomic Data Sequencing Workflow: An example threat model implementation for genomic data sequencing and analysis. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Cybersecurity White Paper (CSWP) NIST CSWP 35 ipd. https://doi.org/10.6028/NIST.CSWP.35.ipd

[8] Garfinkel S, Guttman B, Near J, Dajani A, Singer P (2023) De-identifying Government Datasets: Techniques and Governance. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication 800-188. https://doi.org/10.6028/NIST.SP.800-188

[9] Brooks S, Garcia M, Lefkovitz N, Lightman S, Nadeau E (2017) An Introduction to Privacy Engineering and Risk Management in Federal Systems. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Interagency or Internal Report (IR) NIST IR 8062. https://doi.org/10.6028/NIST.IR.8062

[10] OECD, Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data, OECD/LEGAL/0188. Available at https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0188

[11] Pulivarti R, Martin N, Byers F, Wagner J, Maragh S, Wilson K, Wojtyniak M, Kreider B, Frances A, Edwards S, Morris T, Sheldon J, Ross S, Whitlow P (2023) Cybersecurity of Genomic Data. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Interagency or Internal Report (IR) NIST IR 8432. https://doi.org/10.6028/NIST.IR.8432

[12] Grace T, Nolan T, Burke K, Dudley R, White G, Good T (2006) Guide to Test, Training, and Exercise Programs for IT Plans and Capabilities. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 800-84. https://doi.org/10.6028/NIST.SP.800-84

# Appendix C Threat Modeling Approach

https://pages.nist.gov/nccoe-genomic-data-threat-modeling/Vol_C/Appendix/appendixC.html

# Appendix D Methodology Overview

https://pages.nist.gov/nccoe-genomic-data-threat-modeling/Vol_C/Appendix/appendixD.html

# Appendix E System Description

https://pages.nist.gov/nccoe-genomic-data-threat-modeling/Vol_C/Appendix/appendixE.html

# Appendix F Dataflow Analysis

https://pages.nist.gov/nccoe-genomic-data-threat-modeling/Vol_C/Appendix/appendixF.html

# Appendix G Threat Validation and Prioritization

https://pages.nist.gov/nccoe-genomic-data-threat-modeling/Vol_C/Appendix/appendixG.html