NCCoE Genomic Data Project Update

Genomic Data Profile and Cybersecurity Threat Modeling Workflow White Paper

Ron Pulivarti, Justin Wagner – Principal Investigators (PIs)

Brett Kreider, Julie Snyder, and Kevin Wilson

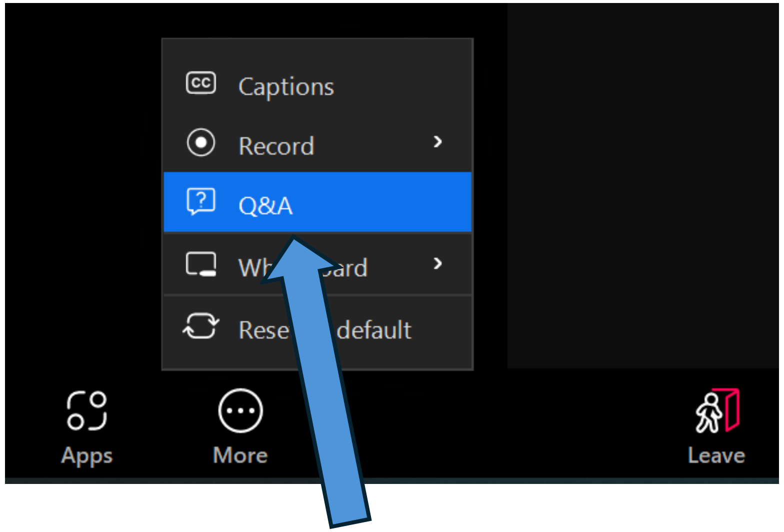
January 13, 2025

This webinar is being recorded

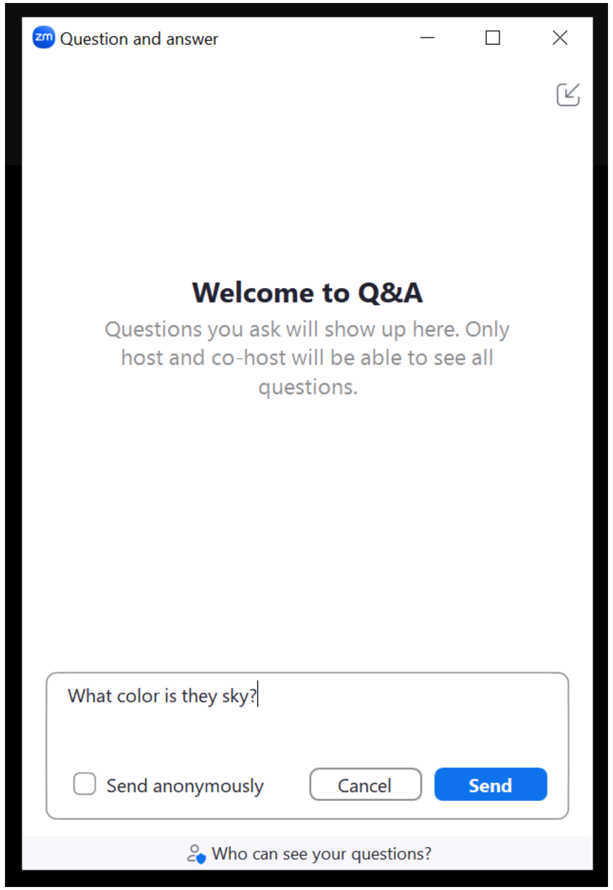
Submitting Questions

Please use the Q&A function to enter your questions.

We will do our best to answer all questions during the Q&A session at the end of this event.



1. To open the Q&A function, click on “More” at the bottom of your screen and select the “Q&A” option.



Welcome to Q&A

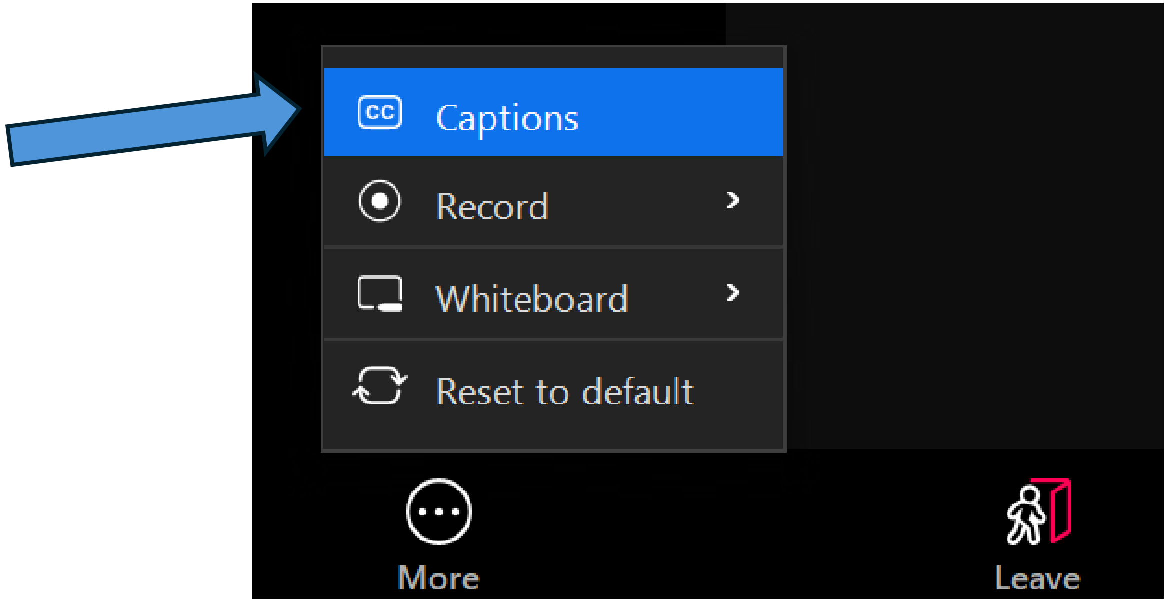
Questions you ask will show up here. Only host and co-host will be able to see all questions.

What color is they sky?

2. Type your question in the text box and click Send

Captions

To open enable captioning during the event, click on "More" at the bottom of your screen and select the "Captions" option.



Speakers

Ron Pulivarti
Genomics Co-PI
NIST NCCoE

Brett Kreider
Project Lead
NCCoE/MITRE

Julie Snyder
Privacy Lead
NCCoE/MITRE

Justin Wagner
Genomics Co-PI
NIST NCCoE

Kevin Wilson, Ph.D.
Cybersecurity Researcher
NCCoE/MITRE

Agenda

- Genomic Data Cybersecurity and Privacy Project Overview
- Genomic Data Cybersecurity and Privacy Frameworks Community Profile (NIST IR 8467)
- Cybersecurity Threat Modeling the Genomic Data Sequencing Workflow (CSWP 35)
- Request for comments and looking ahead

The National Cybersecurity Center of Excellence

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Collaborative hub across governments, industry, and academia to address **real-world cybersecurity challenges.**

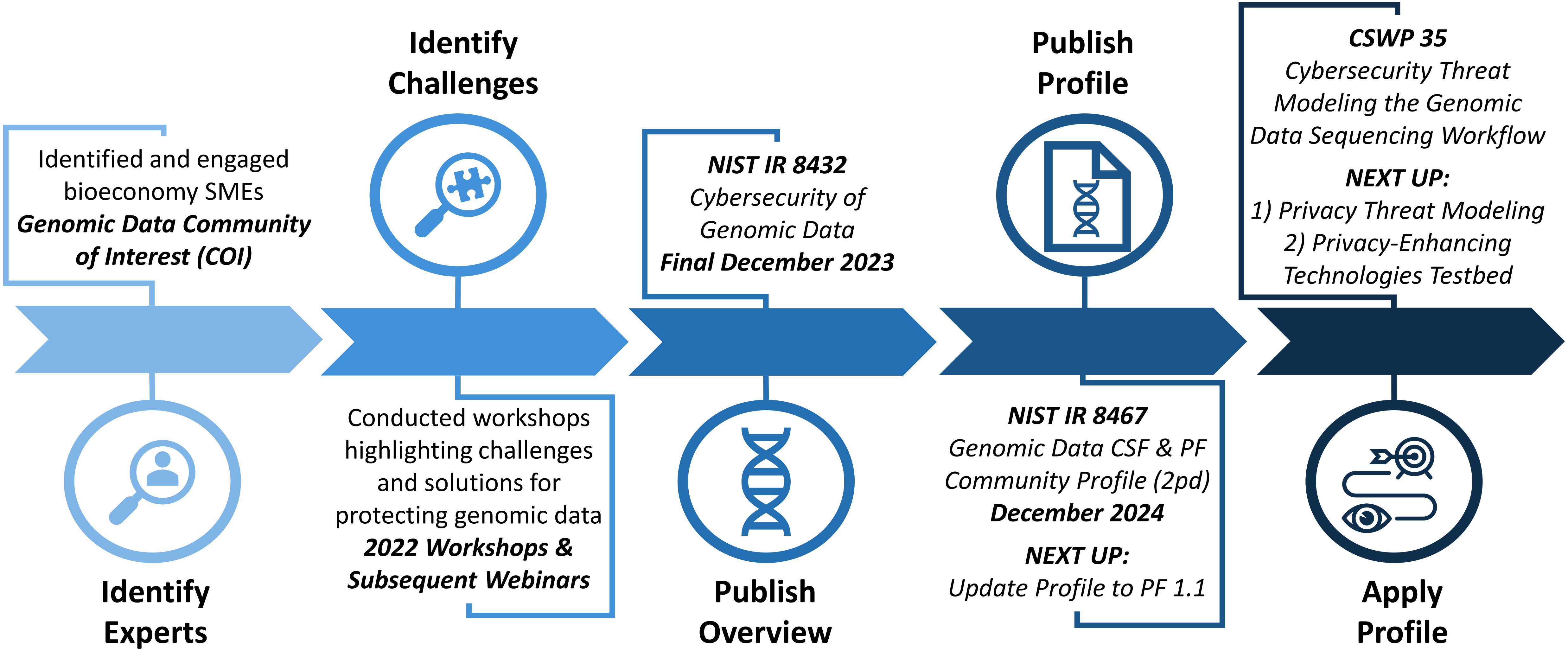
Our Mission: Accelerate Adoption of Secure Technologies

Genomic Data Cybersecurity and Privacy Project Overview

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

nccoe.nist.gov

Genomic Data Project Roadmap



Identified and engaged bioeconomy SMEs

Genomic Data Community of Interest (COI)

Identify Experts

Identify Challenges

Conducted workshops highlighting challenges and solutions for protecting genomic data

2022 Workshops & Subsequent Webinars

NIST IR 8432

Cybersecurity of Genomic Data

Final December 2023

Publish Overview

Publish Profile

NIST IR 8467

Genomic Data CSF & PF Community Profile (2pd)

December 2024

NEXT UP:

Update Profile to PF 1.1

CSWP 35

Cybersecurity Threat Modeling the Genomic Data Sequencing Workflow

NEXT UP:

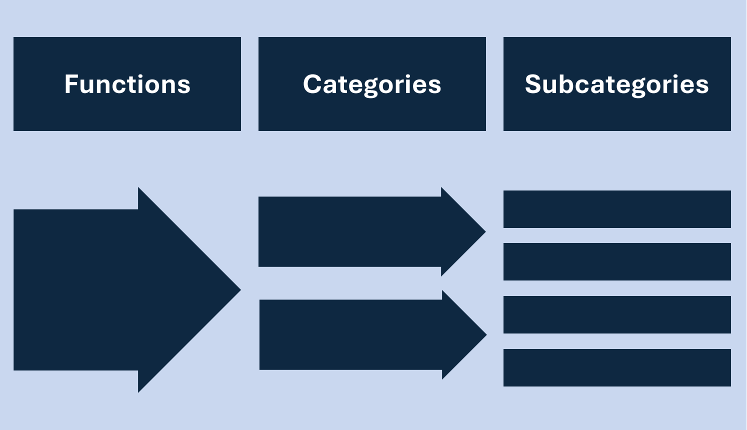
1) Privacy Threat Modeling

2) Privacy-Enhancing Technologies Testbed

Apply Profile

NIST Cybersecurity Framework (CSF) and Privacy Framework (PF) Background

NIST Frameworks Overview

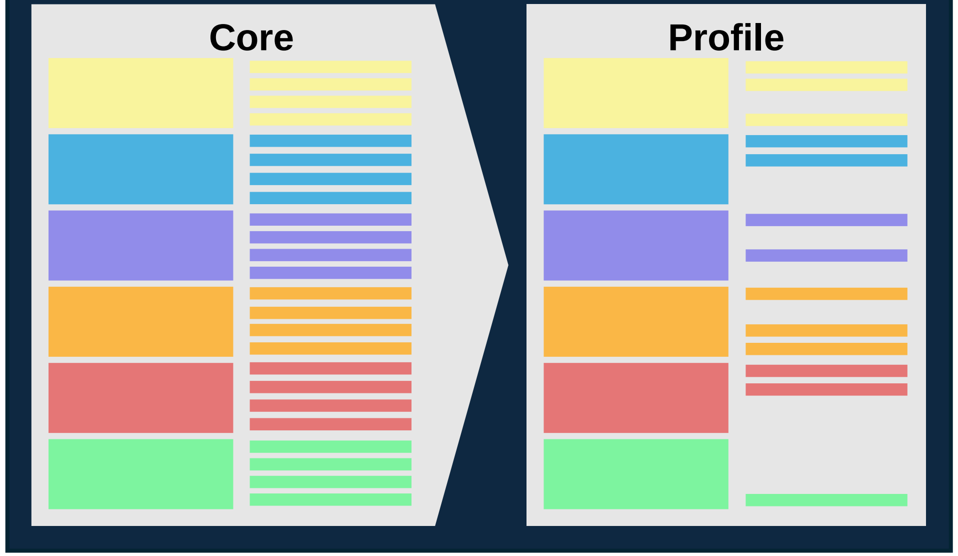


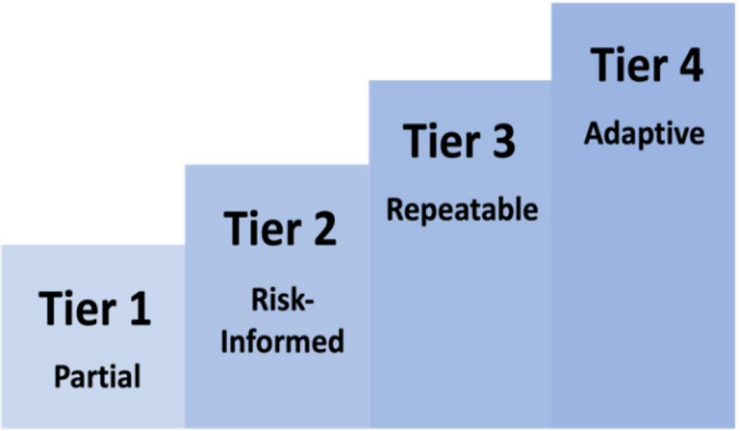
Core

The Core is a set of cybersecurity or privacy outcomes arranged by Function, then Category, and finally Subcategory. While many cybersecurity and privacy risk management activities focus on preventing negative events from occurring, they may also support taking advantage of positive opportunities.

Framework Profiles

Profiles provide a way to understand, tailor, assess, prioritize, and communicate the Core's outcomes based on mission objectives, stakeholder expectations, threat landscape, and requirements.





Tiers

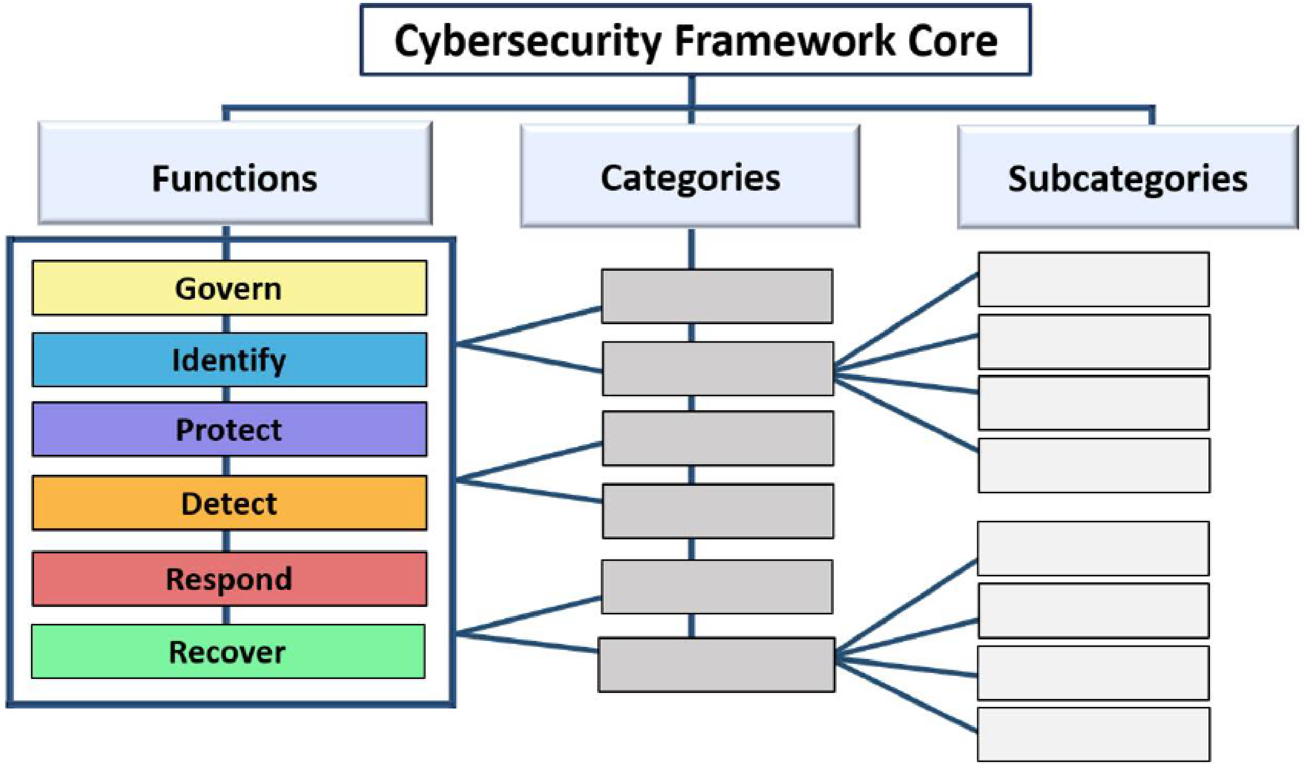
Tiers characterize the rigor of an organization's cybersecurity or privacy risk governance and management practices, and they provide context for how an organization views cybersecurity and privacy risks and the processes in place to manage those risks.

https://doi.org/10.6028/NIST.CSWP.29 and https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.01162020.pdf

NIST Cybersecurity Framework 2.0 Core Overview

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

- **GOVERN** — The organization's cybersecurity risk management strategy, expectations, and policy are established, communicated, and monitored
- **IDENTIFY** — The organization's current cybersecurity risks are understood
- **PROTECT** — Safeguards to manage the organization's cybersecurity risks are used
- **DETECT** — Possible cybersecurity attacks and compromises are found and analyzed
- **RESPOND** — Actions regarding a detected cybersecurity incident are taken
- **RECOVER** — Assets and operations affected by a cybersecurity incident are restored



https://doi.org/10.6028/NIST.CSWP.29

NIST Privacy Framework 1.0 Core Overview

- **Identify-P** — Develop the organizational understanding to manage privacy risk for individuals arising from data processing
- **Govern-P** — Develop and implement the organizational governance structure to enable an ongoing understanding of the organization's risk management priorities that are informed by privacy risk
- **Control-P** — Develop and implement appropriate activities to enable organizations or individuals to manage data with sufficient granularity to manage privacy risks
- **Communicate-P** — Develop and implement appropriate activities to enable organizations and individuals to have a reliable understanding and engage in a dialogue about how data are processed and associated privacy risks
- **Protect-P** — Develop and implement appropriate data processing safeguards

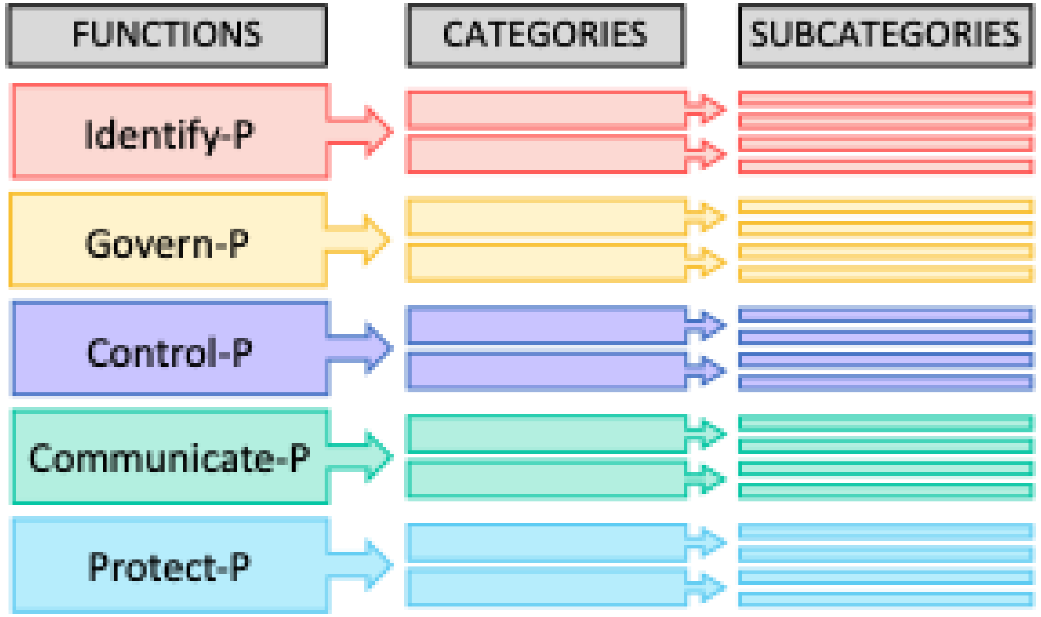


Figure 4: Privacy Framework Core Structure

https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.01162020.pdf

Profiles Overview

NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Framework Profiles

Provide a way to understand, tailor, assess, prioritize, and communicate the Core's outcomes based on mission objectives, stakeholder expectations, threat landscape, and requirements.

Organizational Profiles

Describes an organization's current and/or target posture in terms of the Core's outcomes

Community Profiles

Describes outcomes to address shared interests and goals among multiple organizations

Community Profiles facilitate CSF and PF implementation and help each community address its specific cybersecurity and privacy challenges.

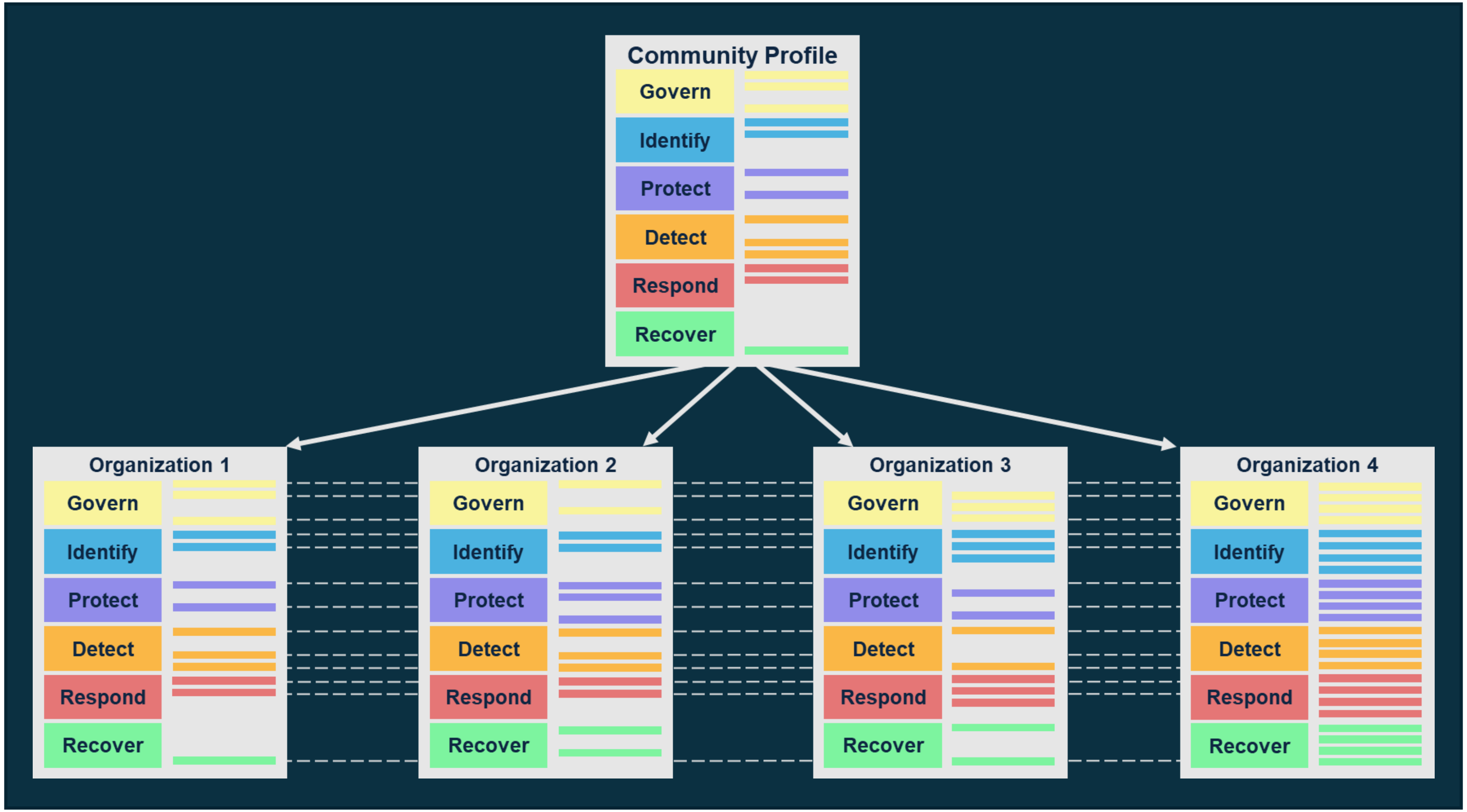
How Organizations Can Use Community Profile Content

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

A Community Profile offers a common view.

Organizations can describe where and why they deviate from the Community Profile and each other and engage in risk discussions with a common starting point.

The NCCoE Resources for Applying NIST Frameworks page is available at: https://www.nccoe.nist.gov/applying-frameworks-resources



Community Profiles Inform Organizational Target Profiles

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE

Community Priorities

CSF and/or PF Core	Aligned Priorities & Outcomes				
	1	2	3	4	5
🟨	⬆	⬇	●	●	●
🟦	●	●	●	⬆	⬇
🟪	●	⬆	●	⬇	●
🟧	⬆	⬆	●	●	⬇
🟥	⬆	●	⬇	●	●
🟩	⬇	●	⬆	●	⬆

Notional Priority Levels*:

⬆ Higher

● Same as Community Profile

⬇ Lower

*Communities and organizations can choose any prioritization schema that works best for them

For more information on Organizational Current Profiles and Target Profiles, see:

- Creating and Using Organizational Profiles Quick Start Guide: *https://doi.org/10.6028/NIST.SP.1301*
- Organizational Profile Template: https://www.nist.gov/document/csf-20-notional-organizational-profile-template

Genomic Data Cybersecurity and Privacy Frameworks Community Profile (NIST IR 8467)

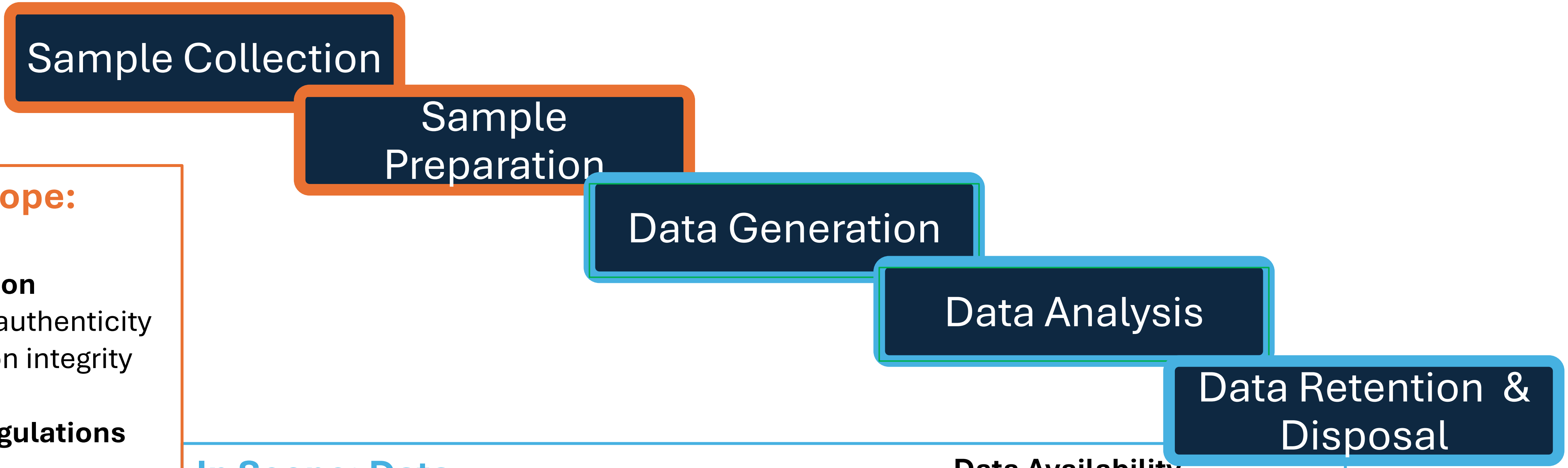
Genomic Data Profile Contents

Executive Summary

1. Introduction: Purpose, Scope, Audience, Structure
2. Overview of Genomic Data
    - 2.1 Ecosystem, Bioeconomy
    - 2.2 Cybersecurity and Privacy Risk Relationship
    - 2.3 Genomic Data Security and Privacy Concerns
3. NIST CSF and PF
4. Genomic Data Profile Development Methodology
5. Genomic Data Mission Objectives
6. Priority CSF and PF Subcategories by Mission Objective

+ Genomic Data Profile Spreadsheet

Project Scope



Not in Scope: Sample

Data Creation

- Sample authenticity
- Extraction integrity

Policy & Regulations

In Scope: Data

Privacy

- Choice
- Consent
- Appropriate use

Data Storage

- Cloud computing
- Breach recovery
- Data volume

Confidentiality

- Deidentification
- Long term sensitivity
- Associational

Data Integrity

- Backup/recovery
- Version control
- Error detection

Data Availability

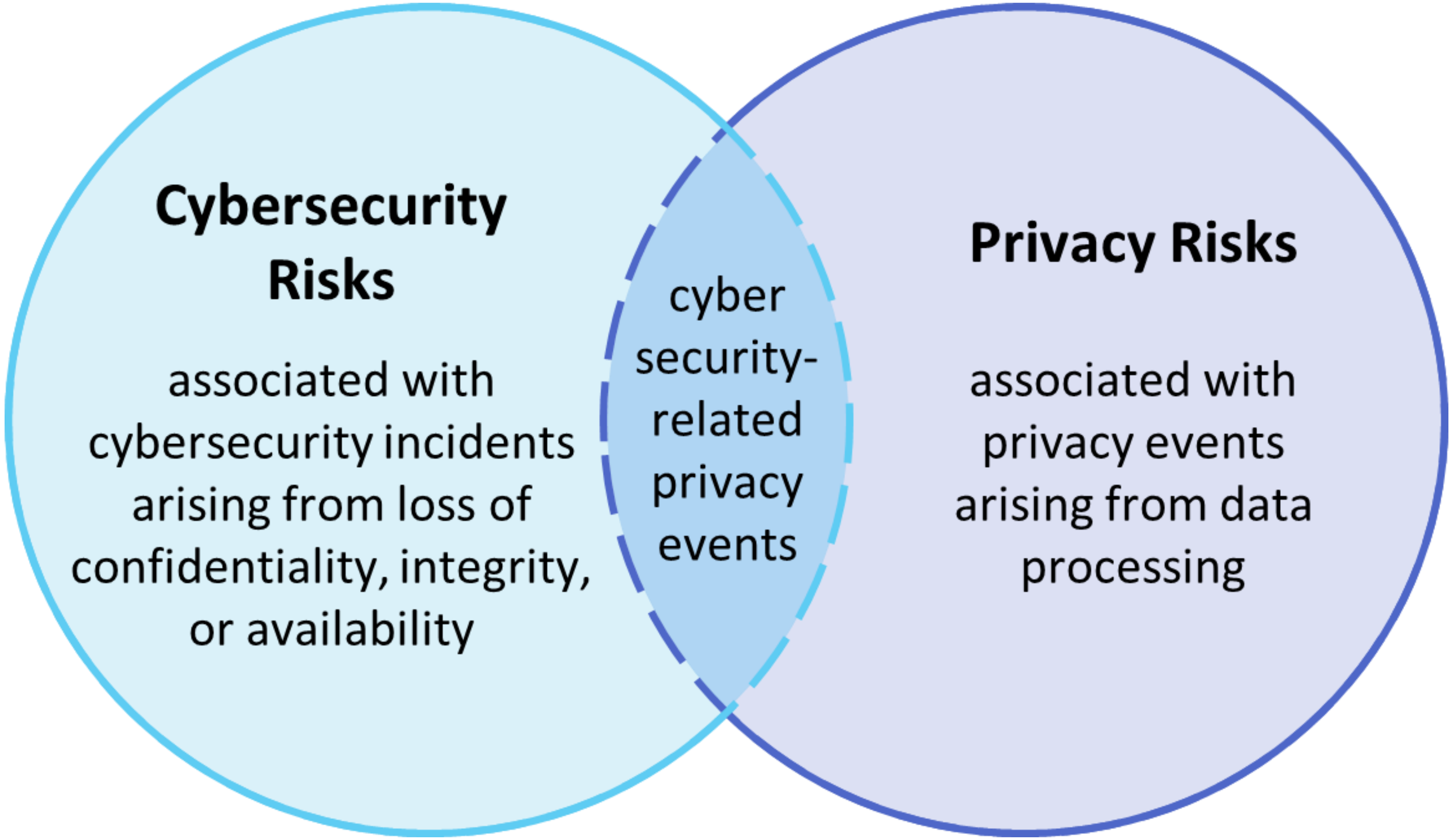
- Authorization/ authentication
- Monitoring and auditing
- Data sharing

Data Retention & Disposal

- Following requirements & agreements
- Device disposal

Cybersecurity and Privacy Risks

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE



Cybersecurity Risks

associated with cybersecurity incidents arising from loss of confidentiality, integrity, or availability

cyber security-related privacy events

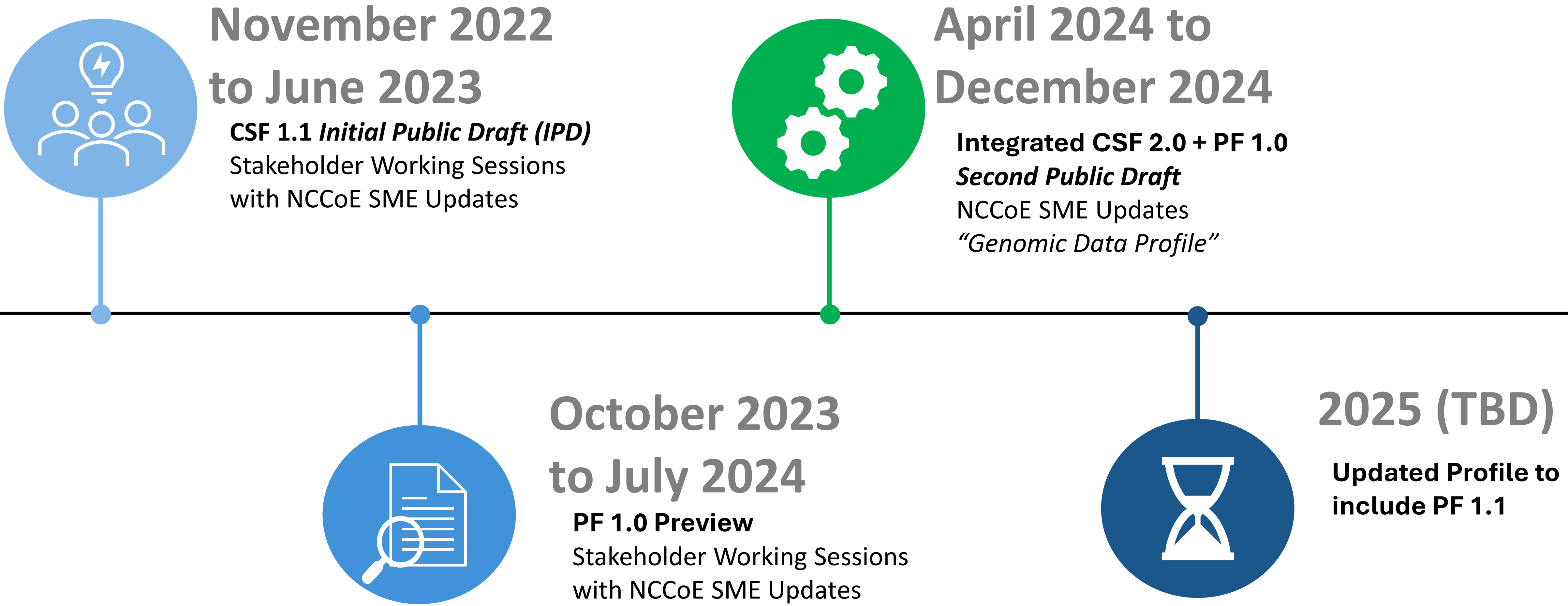
Privacy Risks

associated with privacy events arising from data processing

Cybersecurity and Privacy Risk Relationship (from the NIST Privacy Framework)

Genomic Data Profile Development Approach

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE



November 2022 to June 2023

CSF 1.1 *Initial Public Draft (IPD)*
Stakeholder Working Sessions
with NCCoE SME Updates

October 2023 to July 2024

PF 1.0 Preview
Stakeholder Working Sessions
with NCCoE SME Updates

April 2024 to December 2024

Integrated CSF 2.0 + PF 1.0
Second Public Draft
NCCoE SME Updates
"Genomic Data Profile"

2025 (TBD)

Updated Profile to include PF 1.1

Genomic Data Mission Objectives

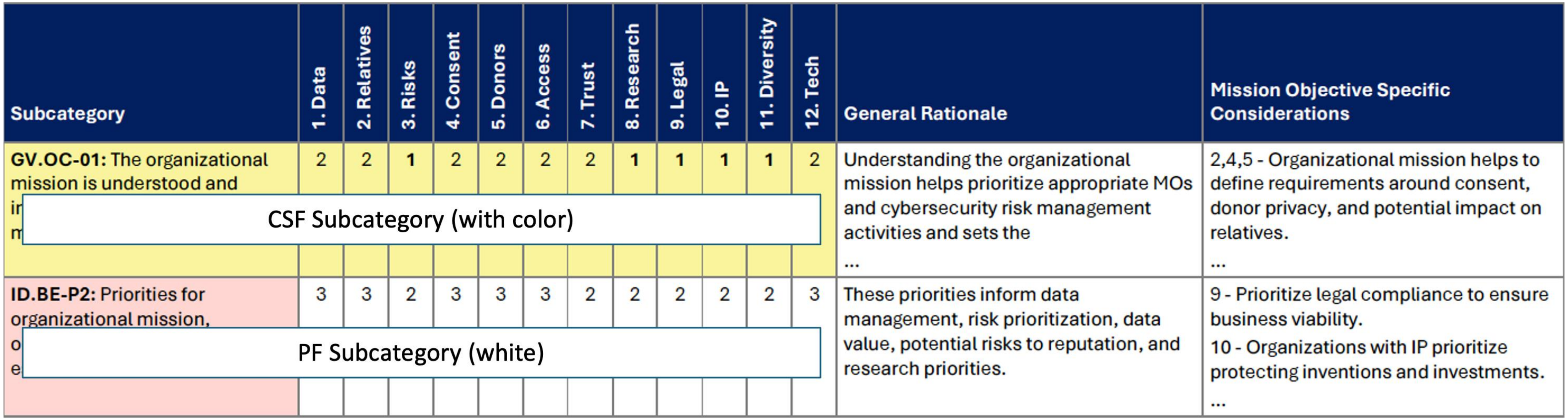
Priority	Mission Objective (Keyword)
1	Manage provenance and data quality throughout the genomic data life cycle (**Data**)
2	Manage privacy risk to existing and future relatives (**Relatives**)
3	Identify, model, and address cybersecurity and privacy risks of processing genomic data (**Risks**)
4	Manage informed consent throughout the genomic data life cycle (**Consent**)
5	Manage privacy risk to donors (**Donors**)
6	Manage authorized data access (**Access**)
7	Maintain trust and manage reputational risk (**Trust**)
8	Facilitate research and education to advance science and technology (**Research**)
9	Maintain compliance with laws and regulations (**Legal**)
10	Protect intellectual property (**IP**)
11	Ensure the degree of diversity is appropriate for processing purposes (**Diversity**)
12	Promote the use of privacy-enhancing technologies as well as secure technologies for sharing genomic data (**Tech**)

These Mission Objectives apply to the integrated Genomic Data Profile.
The CSF addresses cybersecurity aspects of these Mission Objectives.
The PF addresses the privacy aspects of these same Mission Objectives.

Subcategory Prioritization Example

Each table summarizes the Profile information for one CSF Category

12 Mission Objectives



Subcategory	1. Data	2. Relatives	3. Risks	4. Consent	5. Donors	6. Access	7. Trust	8. Research	9. Legal	10. IP	11. Diversity	12. Tech	General Rationale	Mission Objective Specific Considerations
GV.OC-01: The organizational mission is understood and ir m	2	2	**1**	2	2	2	2	**1**	**1**	**1**	**1**	2	Understanding the organizational mission helps prioritize appropriate MOs and cybersecurity risk management activities and sets the …	2,4,5 - Organizational mission helps to define requirements around consent, donor privacy, and potential impact on relatives. …
ID.BE-P2: Priorities for organizational mission, o e	3	3	2	3	3	3	2	2	2	2	2	3	These priorities inform data management, risk prioritization, data value, potential risks to reputation, and research priorities.	9 - Prioritize legal compliance to ensure business viability. 10 - Organizations with IP prioritize protecting inventions and investments. …

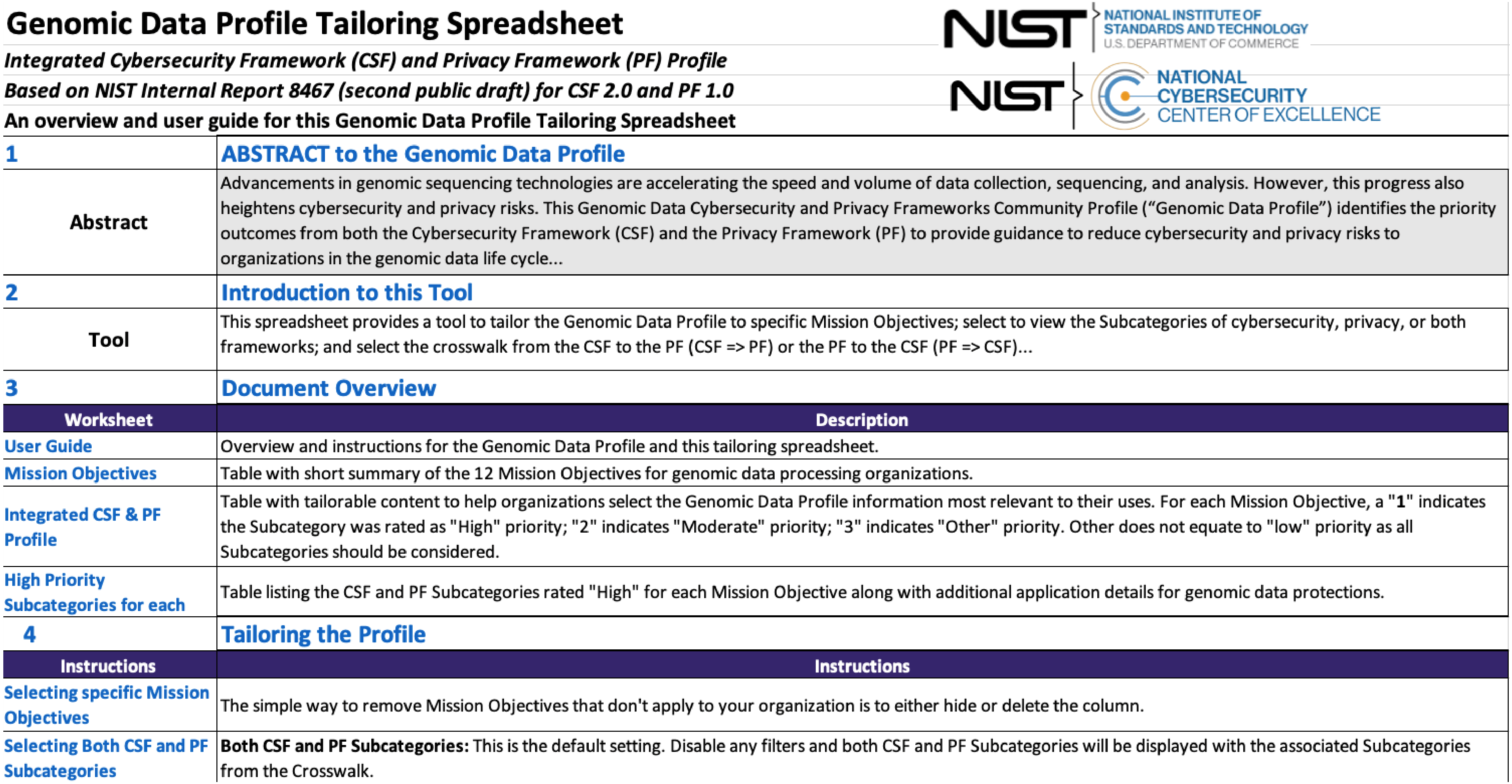
CSF Subcategory (with color)

PF Subcategory (white)

High Priority	Moderate Priority	Other Priority
1	2	3

General Rationale and Mission Objective-Specific Considerations for why a Subcategory should be prioritized for a Mission Objective

Genomic Data Profile Spreadsheet – 1 User Guide Tab



Genomic Data Profile Tailoring Spreadsheet

Integrated Cybersecurity Framework (CSF) and Privacy Framework (PF) Profile

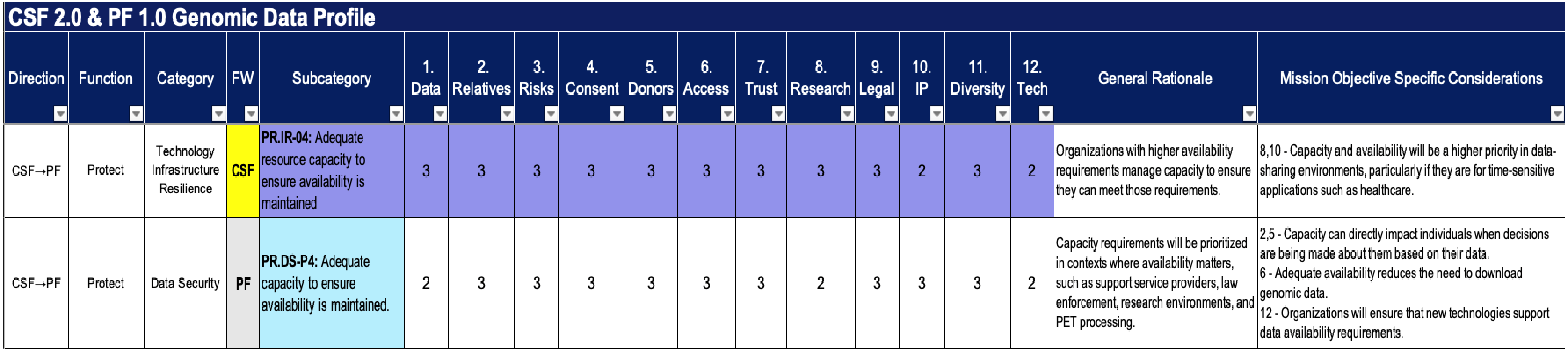
Based on NIST Internal Report 8467 (second public draft) for CSF 2.0 and PF 1.0

An overview and user guide for this Genomic Data Profile Tailoring Spreadsheet

1	**ABSTRACT to the Genomic Data Profile**
Abstract	Advancements in genomic sequencing technologies are accelerating the speed and volume of data collection, sequencing, and analysis. However, this progress also heightens cybersecurity and privacy risks. This Genomic Data Cybersecurity and Privacy Frameworks Community Profile (“Genomic Data Profile”) identifies the priority outcomes from both the Cybersecurity Framework (CSF) and the Privacy Framework (PF) to provide guidance to reduce cybersecurity and privacy risks to organizations in the genomic data life cycle...
2	**Introduction to this Tool**
Tool	This spreadsheet provides a tool to tailor the Genomic Data Profile to specific Mission Objectives; select to view the Subcategories of cybersecurity, privacy, or both frameworks; and select the crosswalk from the CSF to the PF (CSF => PF) or the PF to the CSF (PF => CSF)...
3	**Document Overview**
Worksheet	**Description**
User Guide	Overview and instructions for the Genomic Data Profile and this tailoring spreadsheet.
Mission Objectives	Table with short summary of the 12 Mission Objectives for genomic data processing organizations.
Integrated CSF & PF Profile	Table with tailorable content to help organizations select the Genomic Data Profile information most relevant to their uses. For each Mission Objective, a "**1**" indicates the Subcategory was rated as "High" priority; "2" indicates "Moderate" priority; "3" indicates "Other" priority. Other does not equate to "low" priority as all Subcategories should be considered.
High Priority Subcategories for each	Table listing the CSF and PF Subcategories rated "High" for each Mission Objective along with additional application details for genomic data protections.
4	**Tailoring the Profile**
Instructions	**Instructions**
Selecting specific Mission Objectives	The simple way to remove Mission Objectives that don't apply to your organization is to either hide or delete the column.
Selecting Both CSF and PF Subcategories	**Both CSF and PF Subcategories:** This is the default setting. Disable any filters and both CSF and PF Subcategories will be displayed with the associated Subcategories from the Crosswalk.

Genomic Data Profile Spreadsheet – 2
Customizable Integrated Profile

- Select applicable Mission Objectives
- Choose to display CSF, PF, or Both
- Choose crosswalk CSF => PF or PF => CSF

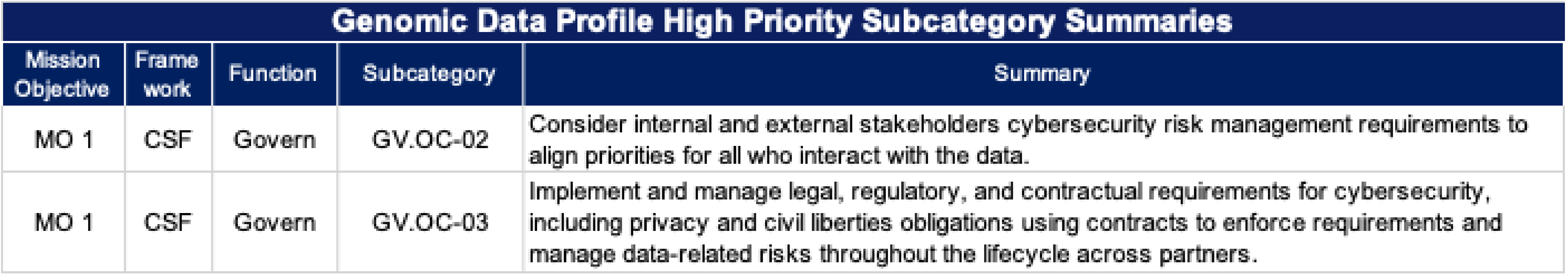


CSF 2.0 & PF 1.0 Genomic Data Profile

Direction	Function	Category	FW	Subcategory	1. Data	2. Relatives	3. Risks	4. Consent	5. Donors	6. Access	7. Trust	8. Research	9. Legal	10. IP	11. Diversity	12. Tech	General Rationale	Mission Objective Specific Considerations
CSF→PF	Protect	Technology Infrastructure Resilience	CSF	**PR.IR-04:** Adequate resource capacity to ensure availability is maintained	3	3	3	3	3	3	3	3	3	2	3	2	Organizations with higher availability requirements manage capacity to ensure they can meet those requirements.	8,10 - Capacity and availability will be a higher priority in data-sharing environments, particularly if they are for time-sensitive applications such as healthcare.
CSF→PF	Protect	Data Security	PF	**PR.DS-P4:** Adequate capacity to ensure availability is maintained.	2	3	3	3	3	3	3	2	3	3	3	2	Capacity requirements will be prioritized in contexts where availability matters, such as support service providers, law enforcement, research environments, and PET processing.	2,5 - Capacity can directly impact individuals when decisions are being made about them based on their data. 6 - Adequate availability reduces the need to download genomic data. 12 - Organizations will ensure that new technologies support data availability requirements.

Genomic Data Profile Spreadsheet – 3 HIGH Priority Subcategories Tab

- Identifies only HIGH Priority Subcategories for each Mission Objective
  - Select High Subcategories from the CSF, PF, or Both
  - Summary provides an overview of the Subcategory with some context from the General Rationale and Mission Objective specific considerations columns



Genomic Data Profile High Priority Subcategory Summaries

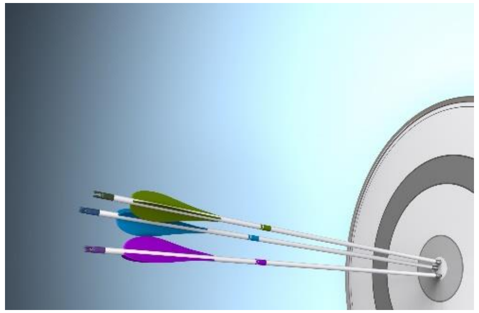
Mission Objective	Frame work	Function	Subcategory	Summary
MO 1	CSF	Govern	GV.OC-02	Consider internal and external stakeholders cybersecurity risk management requirements to align priorities for all who interact with the data.
MO 1	CSF	Govern	GV.OC-03	Implement and manage legal, regulatory, and contractual requirements for cybersecurity, including privacy and civil liberties obligations using contracts to enforce requirements and manage data-related risks throughout the lifecycle across partners.

Benefits and Uses of Community Profiles

Use **shared taxonomy** for cybersecurity and privacy in the context of the community

Align requirements from multiple sources

Leverage expertise across the community



Encourage **common target** outcomes

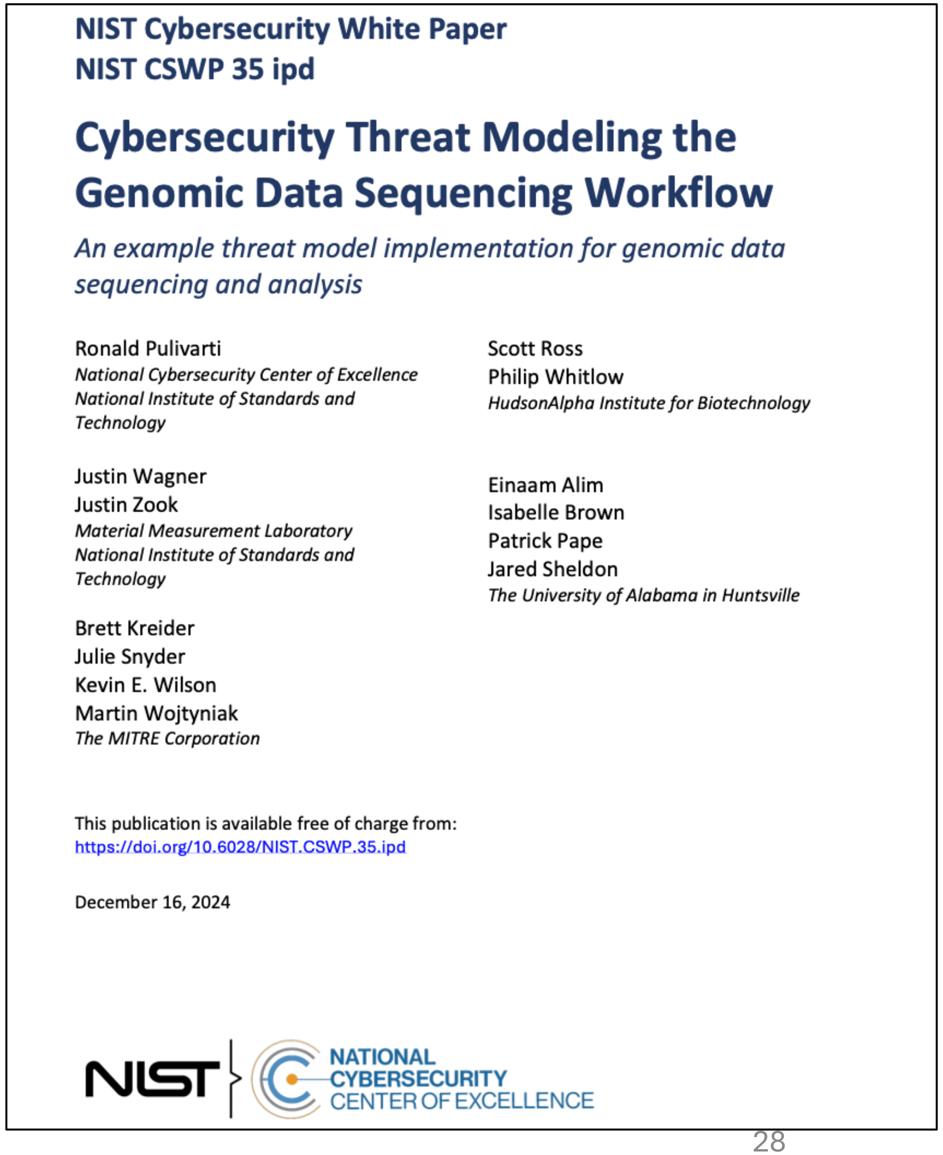
Minimize the burden by working together

Communicate about cybersecurity & privacy risks

Cybersecurity Threat Modeling the Genomic Data Sequencing Workflow (CSWP 35)

NIST Cybersecurity White Paper (CSWP) 35

- *Cybersecurity Threat Modeling the Genomic Data Sequencing Workflow*
- Provides an example of how to conduct threat modeling including
  1. Diagramming dataflows
  2. Identifying threats
  3. Addressing threats
  4. Improving the process
- Uses the Genomic Data Profile, STRIDE, Attack Trees, & MITRE ATT&CK® Techniques, Tactics, & Protocols (TTPs)



NIST Cybersecurity White Paper
NIST CSWP 35 ipd

Cybersecurity Threat Modeling the Genomic Data Sequencing Workflow

An example threat model implementation for genomic data sequencing and analysis

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

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

Brett Kreider
Julie Snyder
Kevin E. Wilson
Martin Wojtyniak
The MITRE Corporation

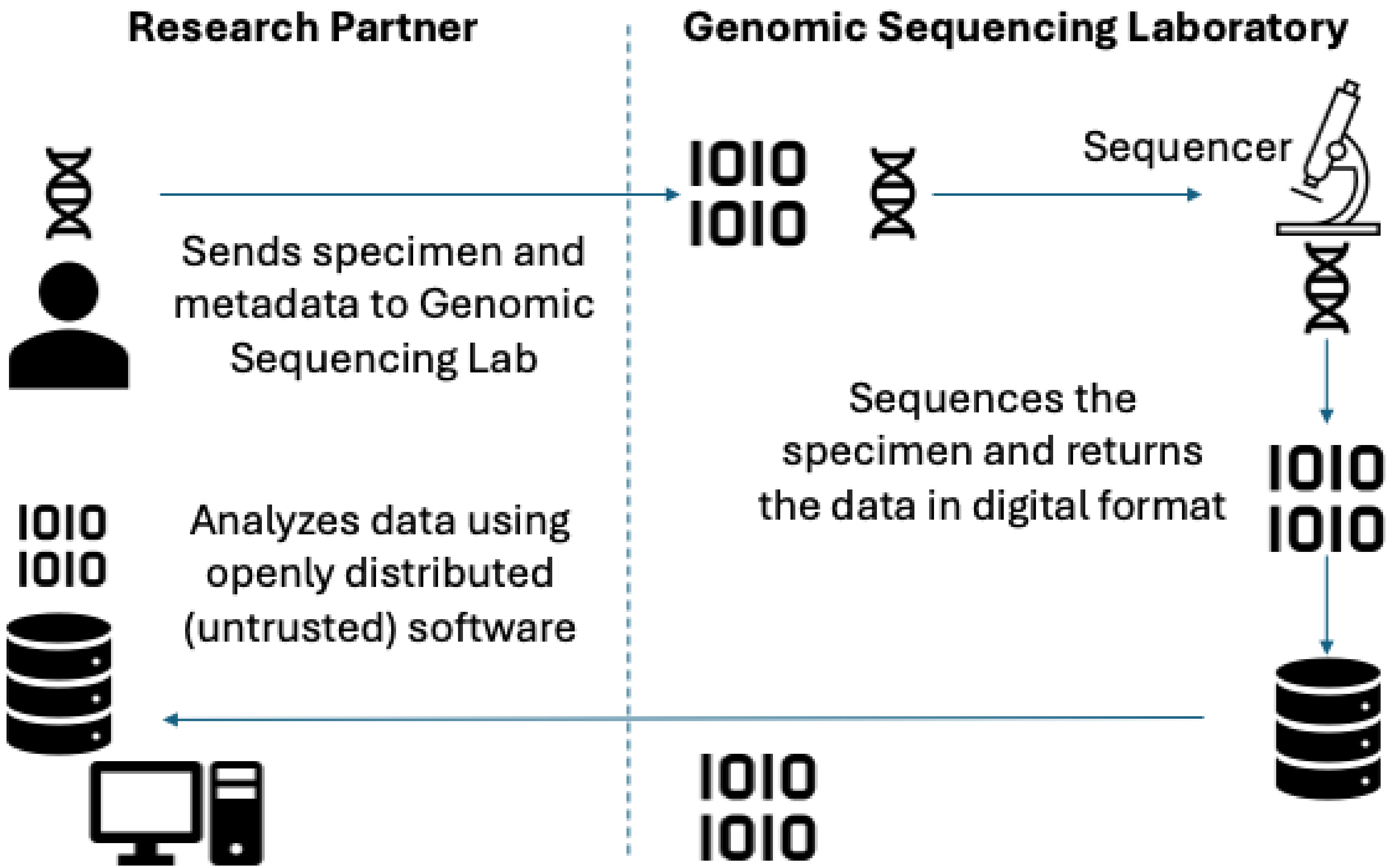
Scott Ross
Philip Whitlow
HudsonAlpha Institute for Biotechnology

Einaam Alim
Isabelle Brown
Patrick Pape
Jared Sheldon
The University of Alabama in Huntsville

This publication is available free of charge from:
https://doi.org/10.6028/NIST.CSWP.35.ipd

December 16, 2024

Use Case: Genomic Sequencing



Research Partner

Genomic Sequencing Laboratory

Sends specimen and metadata to Genomic Sequencing Lab

Sequencer

Sequences the specimen and returns the data in digital format

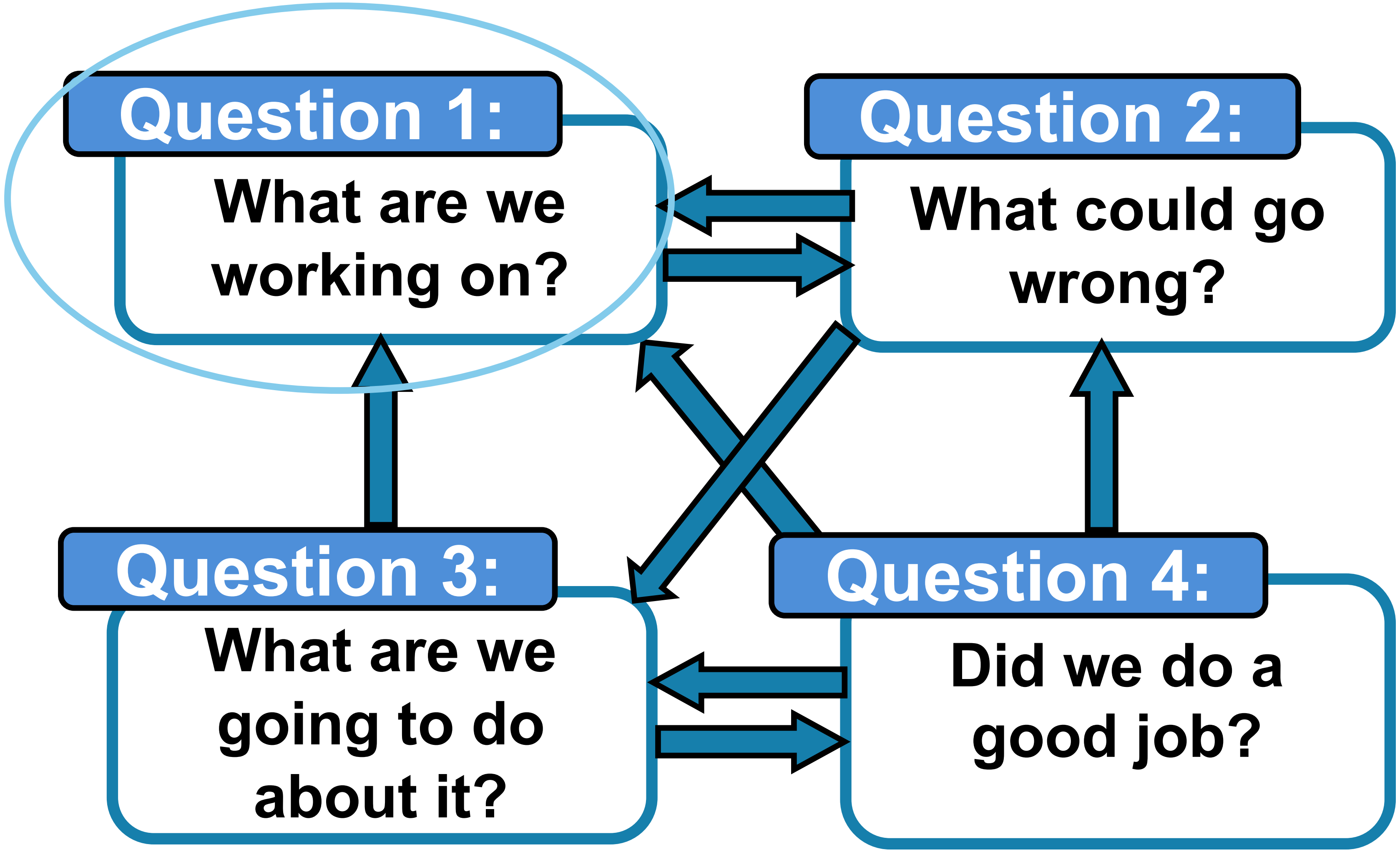
Analyzes data using openly distributed (untrusted) software

Using the Genomic Data Profile

Identify Mission Objectives for the selected Genomic Workflow from the *Genomic Data Cybersecurity and Privacy Frameworks Community Profile*

- Mission Objective 1: Manage provenance and data quality throughout the genomic data lifecycle (**Data**)
- Mission Objective 3: Identify, model, and address cybersecurity and privacy risks of processing genomic data (**Risk**)
- Mission Objective 8: Facilitate research and education to advance science and technology (**Tech**)

Threat Modeling
4-Question Framework Process



Question 1:
What are we working on?

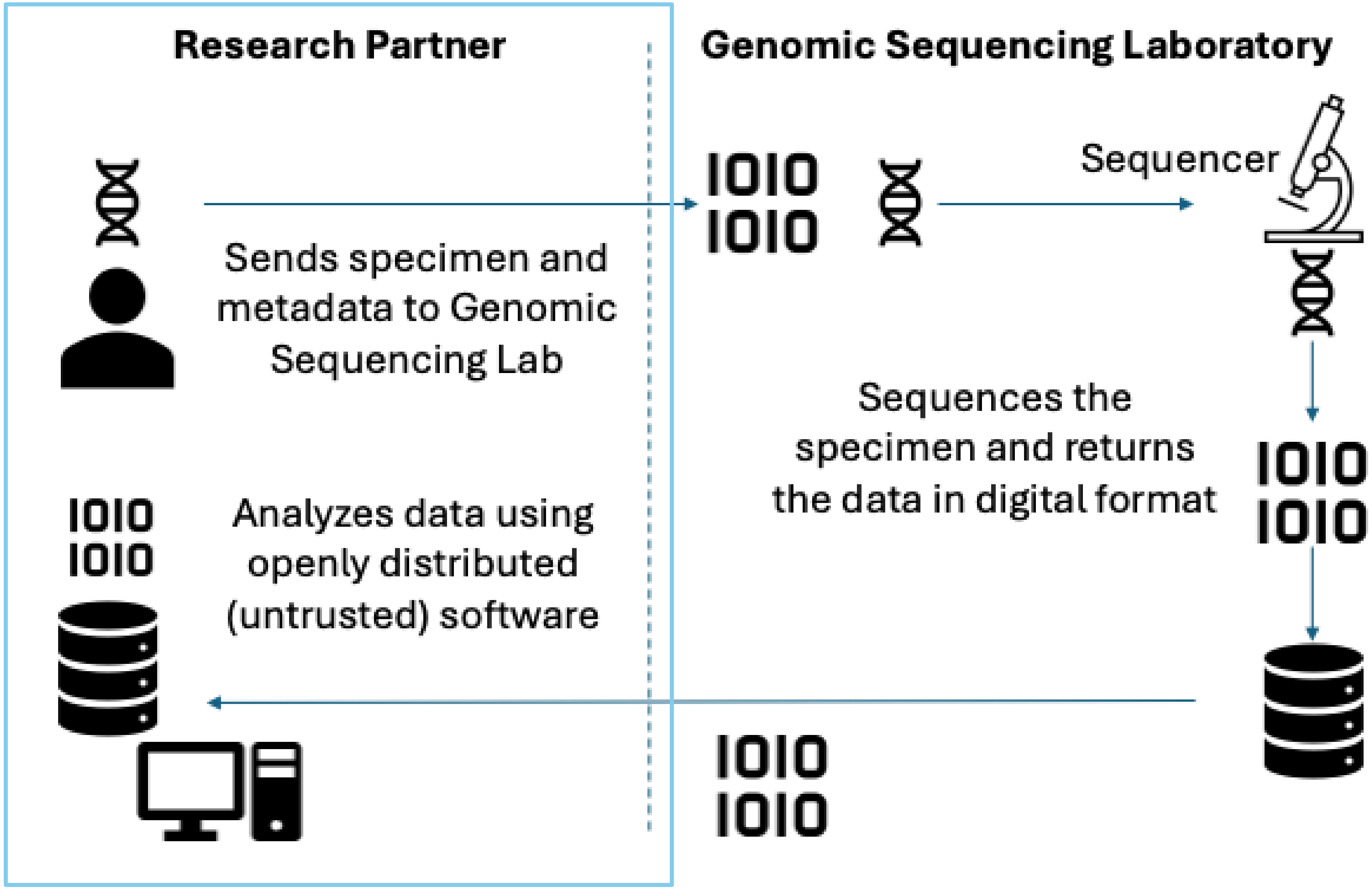
Question 2:
What could go wrong?

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

Question 4:
Did we do a good job?

Threatmodelmanifesto.org

1. What are we working on?



Research Partner

Genomic Sequencing Laboratory

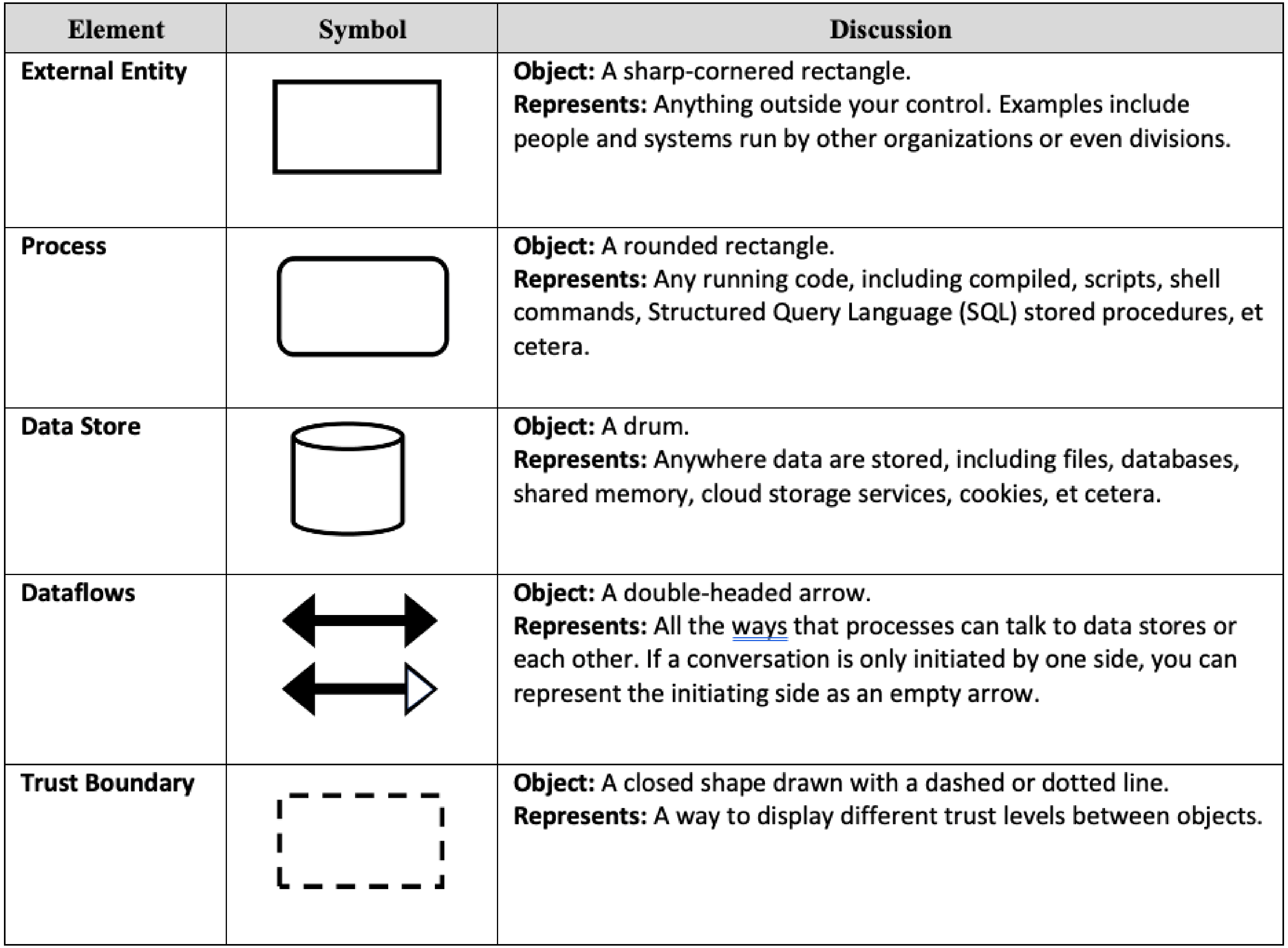
Sends specimen and metadata to Genomic Sequencing Lab

Sequencer

Sequences the specimen and returns the data in digital format

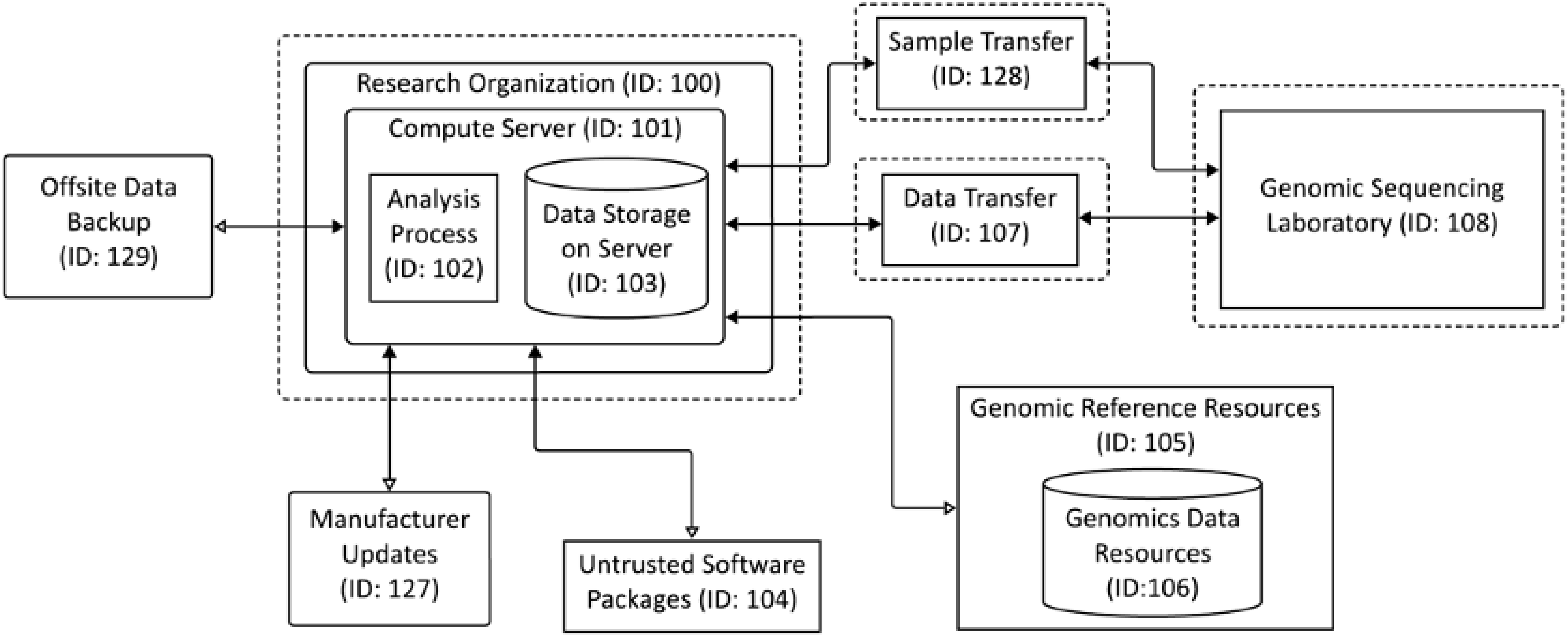
Analyzes data using openly distributed (untrusted) software

Dataflow Diagram (DFD) Legend

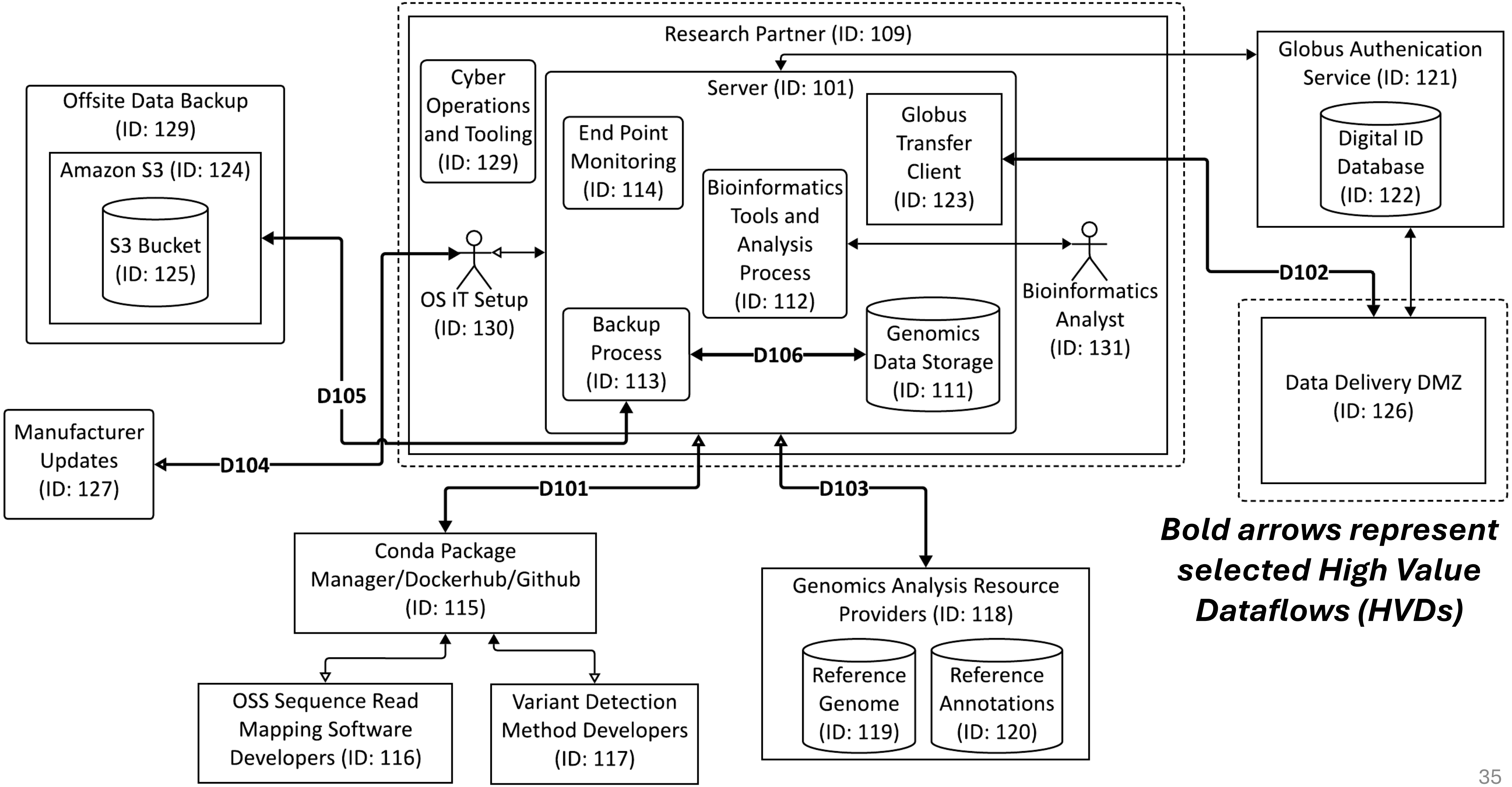


Element	Symbol	Discussion
External Entity	□	**Object:** A sharp-cornered rectangle. **Represents:** Anything outside your control. Examples include people and systems run by other organizations or even divisions.
Process	▢	**Object:** A rounded rectangle. **Represents:** Any running code, including compiled, scripts, shell commands, Structured Query Language (SQL) stored procedures, et cetera.
Data Store	⛁	**Object:** A drum. **Represents:** Anywhere data are stored, including files, databases, shared memory, cloud storage services, cookies, et cetera.
Dataflows	◀━▶ ◀━▷	**Object:** A double-headed arrow. **Represents:** All the ways that processes can talk to data stores or each other. If a conversation is only initiated by one side, you can represent the initiating side as an empty arrow.
Trust Boundary	⬚	**Object:** A closed shape drawn with a dashed or dotted line. **Represents:** A way to display different trust levels between objects.

Research Partner – High Level DFD


Research Organization (ID: 100)
Compute Server (ID: 101)
Analysis Process (ID: 102)
Data Storage on Server (ID: 103)
Offsite Data Backup (ID: 129)
Sample Transfer (ID: 128)
Data Transfer (ID: 107)
Genomic Sequencing Laboratory (ID: 108)
Genomic Reference Resources (ID: 105)
Genomics Data Resources (ID:106)
Manufacturer Updates (ID: 127)
Untrusted Software Packages (ID: 104)

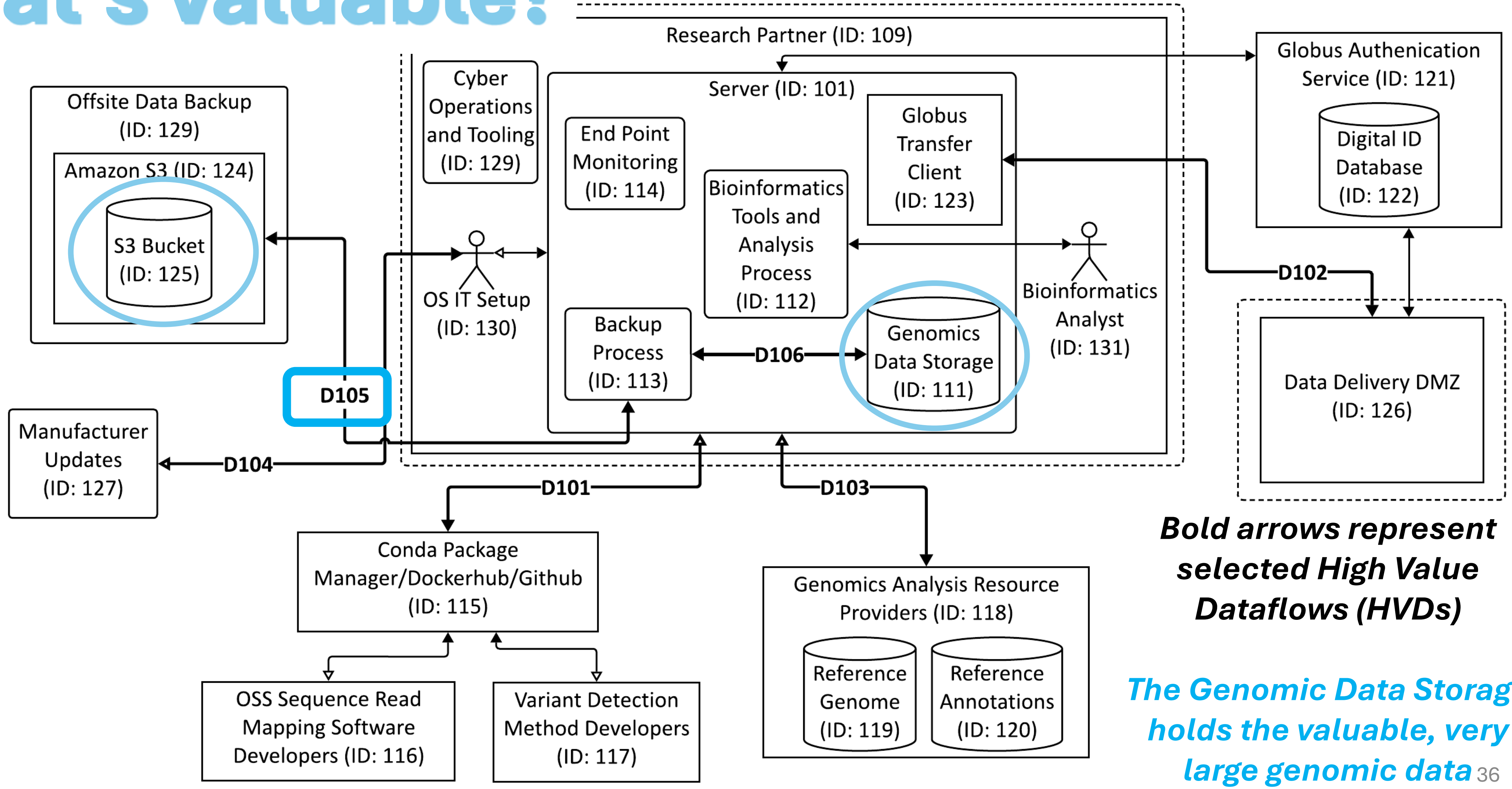
Research Partner – Detailed DFD


Research Partner (ID: 109)
Cyber Operations and Tooling (ID: 129)
Server (ID: 101)
End Point Monitoring (ID: 114)
Globus Transfer Client (ID: 123)
Bioinformatics Tools and Analysis Process (ID: 112)
OS IT Setup (ID: 130)
Bioinformatics Analyst (ID: 131)
Backup Process (ID: 113)
D106
Genomics Data Storage (ID: 111)
Offsite Data Backup (ID: 129)
Amazon S3 (ID: 124)
S3 Bucket (ID: 125)
Globus Authenication Service (ID: 121)
Digital ID Database (ID: 122)
D102
Data Delivery DMZ (ID: 126)
D105
Manufacturer Updates (ID: 127)
D104
D101
D103
Conda Package Manager/Dockerhub/Github (ID: 115)
OSS Sequence Read Mapping Software Developers (ID: 116)
Variant Detection Method Developers (ID: 117)
Genomics Analysis Resource Providers (ID: 118)
Reference Genome (ID: 119)
Reference Annotations (ID: 120)

Bold arrows represent selected High Value Dataflows (HVDs)

Research Partner – Detailed DFD

What's valuable?

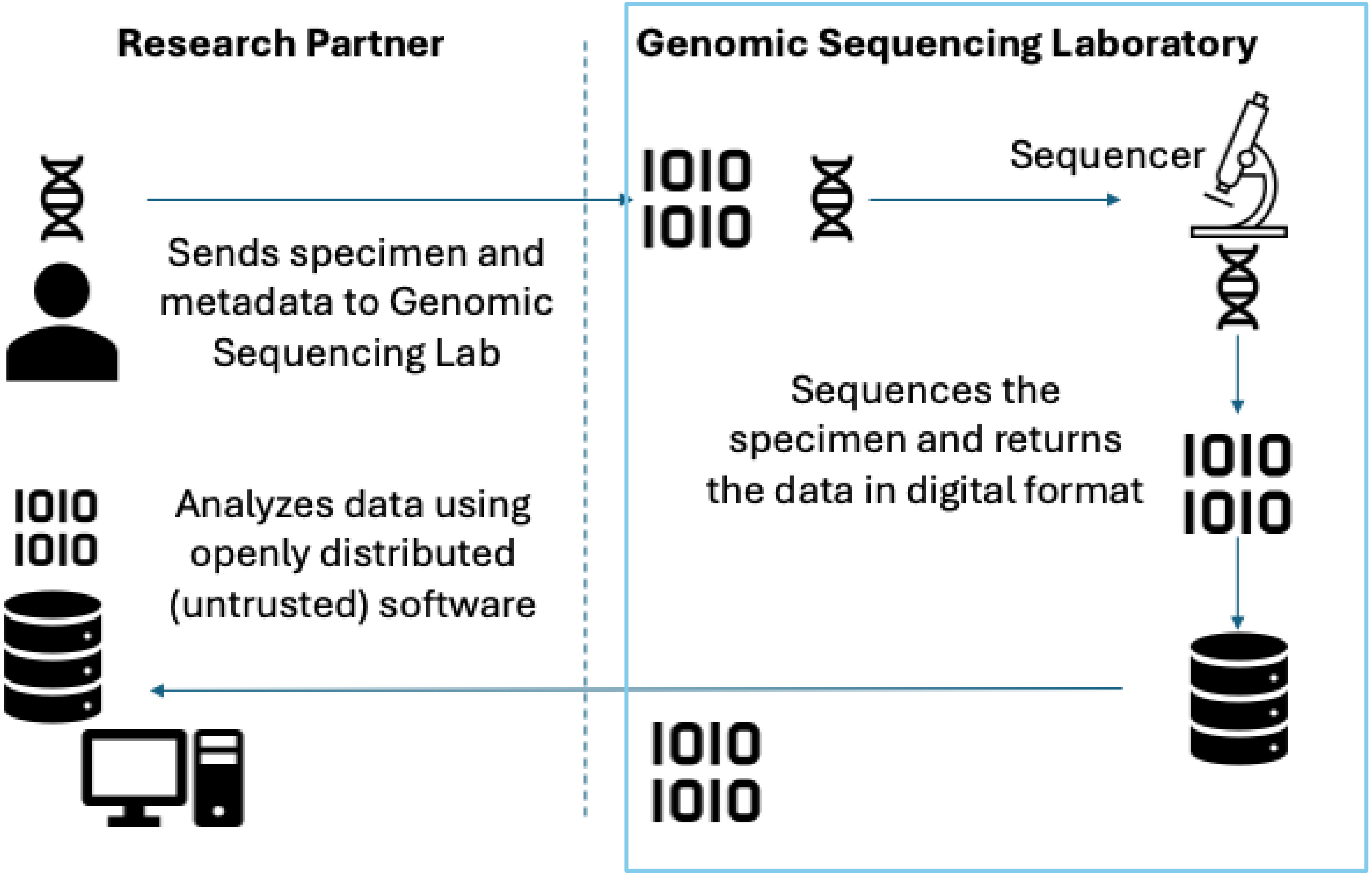


Research Partner (ID: 109)
Server (ID: 101)
Offsite Data Backup (ID: 129)
Amazon S3 (ID: 124)
S3 Bucket (ID: 125)
Cyber Operations and Tooling (ID: 129)
End Point Monitoring (ID: 114)
Globus Transfer Client (ID: 123)
Bioinformatics Tools and Analysis Process (ID: 112)
OS IT Setup (ID: 130)
Bioinformatics Analyst (ID: 131)
Backup Process (ID: 113)
Genomics Data Storage (ID: 111)
Globus Authenication Service (ID: 121)
Digital ID Database (ID: 122)
Data Delivery DMZ (ID: 126)
Manufacturer Updates (ID: 127)
D105
D104
D106
D102
D101
D103
Conda Package Manager/Dockerhub/Github (ID: 115)
OSS Sequence Read Mapping Software Developers (ID: 116)
Variant Detection Method Developers (ID: 117)
Genomics Analysis Resource Providers (ID: 118)
Reference Genome (ID: 119)
Reference Annotations (ID: 120)

Bold arrows represent selected High Value Dataflows (HVDs)

The Genomic Data Storage holds the valuable, very large genomic data

1. What are we working on?



Research Partner

Genomic Sequencing Laboratory

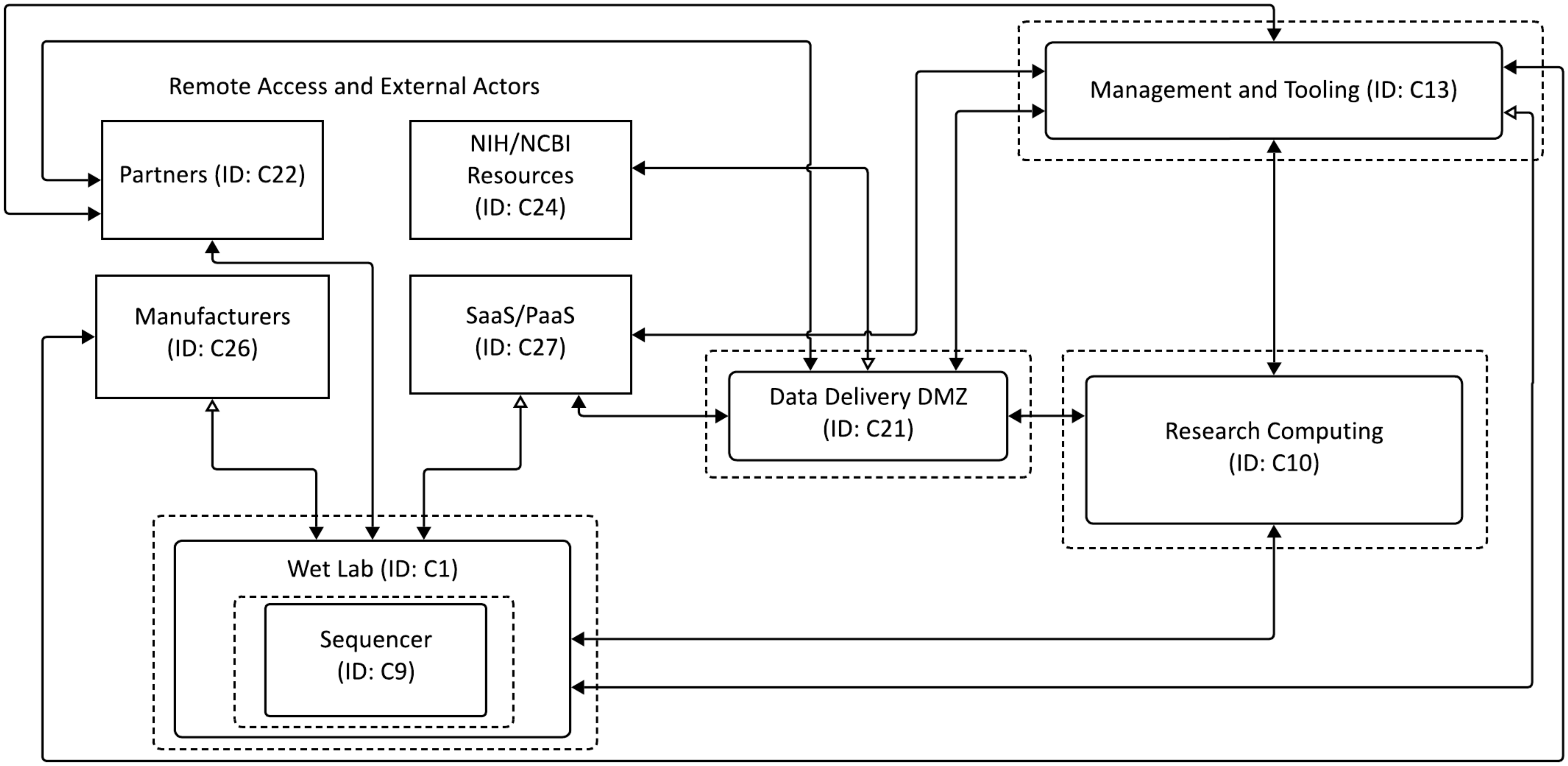
Sends specimen and metadata to Genomic Sequencing Lab

Sequencer

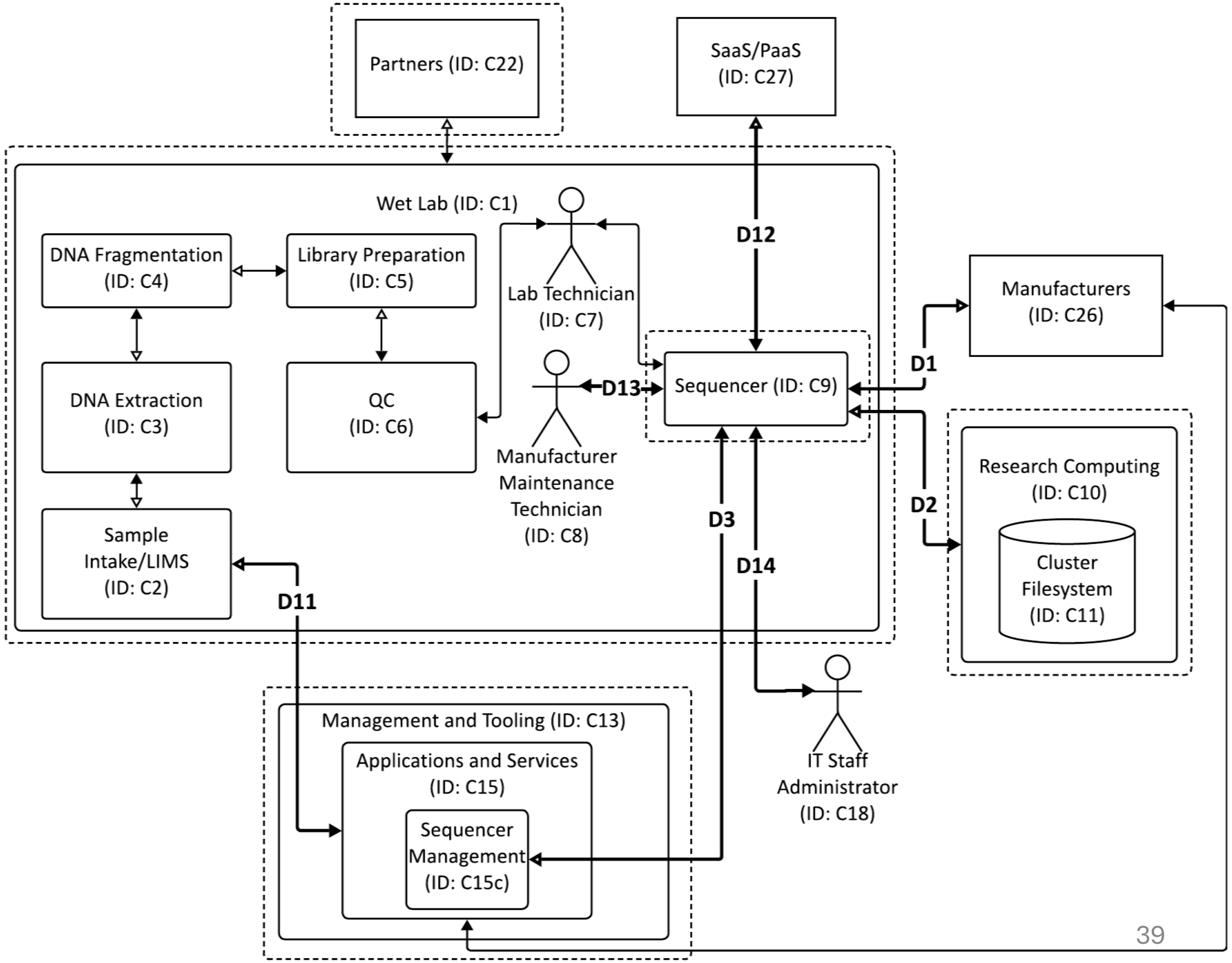
Sequences the specimen and returns the data in digital format

Analyzes data using openly distributed (untrusted) software

Genomic Sequencing Lab High Level DFD


Remote Access and External Actors
Partners (ID: C22)
NIH/NCBI Resources (ID: C24)
Manufacturers (ID: C26)
SaaS/PaaS (ID: C27)
Management and Tooling (ID: C13)
Data Delivery DMZ (ID: C21)
Research Computing (ID: C10)
Wet Lab (ID: C1)
Sequencer (ID: C9)

Wet Lab Detailed DFD



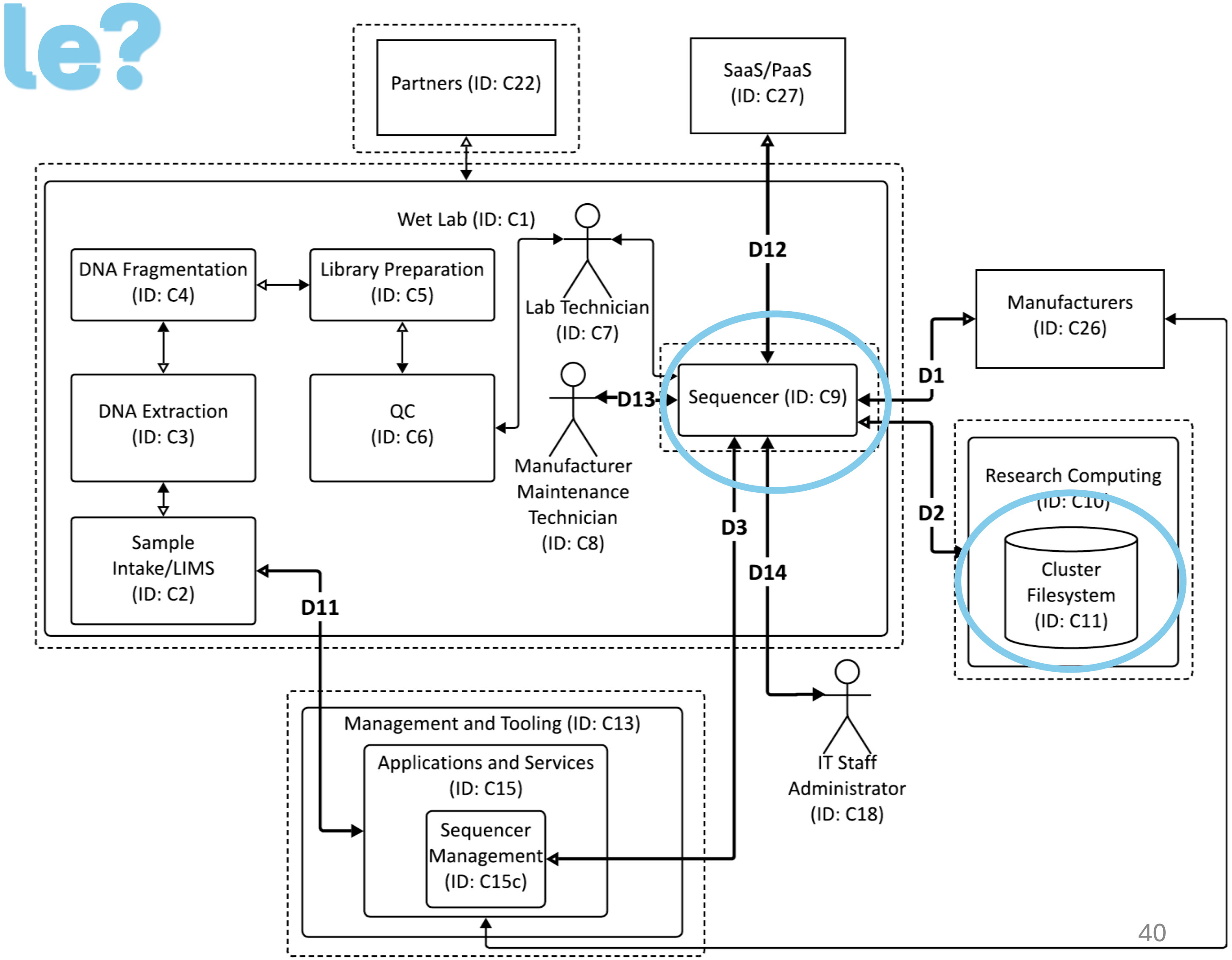
Partners (ID: C22)
SaaS/PaaS (ID: C27)
Wet Lab (ID: C1)
DNA Fragmentation (ID: C4)
Library Preparation (ID: C5)
Lab Technician (ID: C7)
D12
Manufacturers (ID: C26)
D1
DNA Extraction (ID: C3)
QC (ID: C6)
D13
Sequencer (ID: C9)
Manufacturer Maintenance Technician (ID: C8)
Research Computing (ID: C10)
D2
Cluster Filesystem (ID: C11)
D3
D14
Sample Intake/LIMS (ID: C2)
D11
IT Staff Administrator (ID: C18)
Management and Tooling (ID: C13)
Applications and Services (ID: C15)
Sequencer Management (ID: C15c)

Wet Lab Detailed DFD

What's valuable?

1. The Genomic Sequencer is expensive and contains the sample and digital genomic data sequence

2. The Cluster Filesystem stores the digital genomic data sequence


Partners (ID: C22)
SaaS/PaaS (ID: C27)
Wet Lab (ID: C1)
DNA Fragmentation (ID: C4)
Library Preparation (ID: C5)
Lab Technician (ID: C7)
D12
Manufacturers (ID: C26)
D1
DNA Extraction (ID: C3)
QC (ID: C6)
D13
Sequencer (ID: C9)
Manufacturer Maintenance Technician (ID: C8)
D2
Research Computing (ID: C10)
Cluster Filesystem (ID: C11)
D3
D14
Sample Intake/LIMS (ID: C2)
D11
Management and Tooling (ID: C13)
Applications and Services (ID: C15)
Sequencer Management (ID: C15c)
IT Staff Administrator (ID: C18)

Dataflow Diagrams

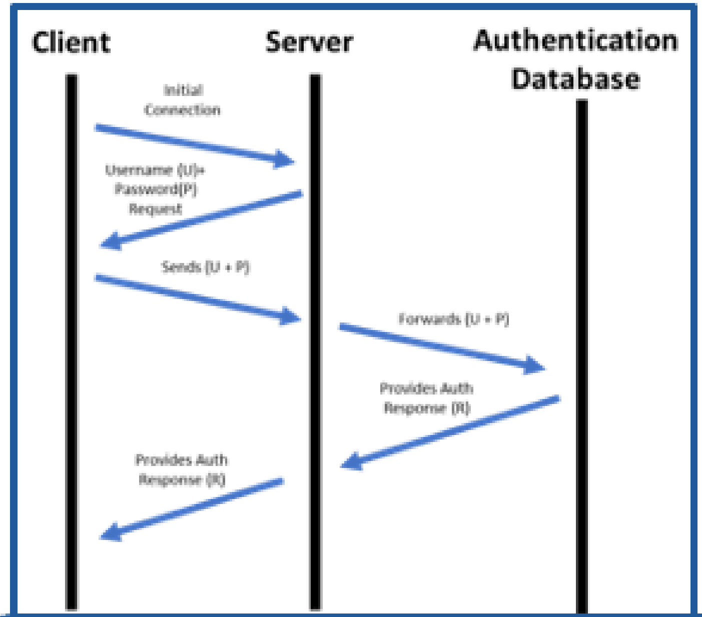
Strengths

- Illustrates dataflows (which systems share data)
- Highlights interfaces
- Defines trust boundaries and areas of control
- Gives basic overview of a system
- Works well with threat identification techniques such as STRIDE

Weaknesses

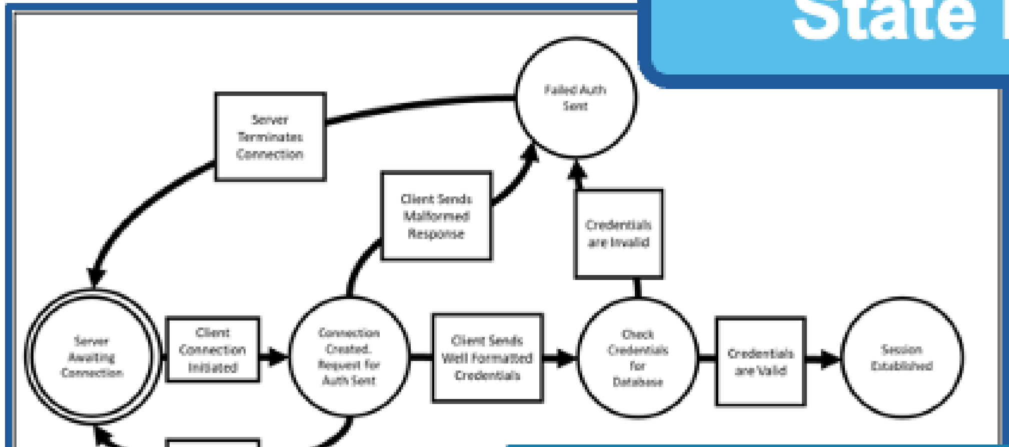
- Does not depict protocols
- Does not describe why systems share data
- Does not cover the actual data being transferred

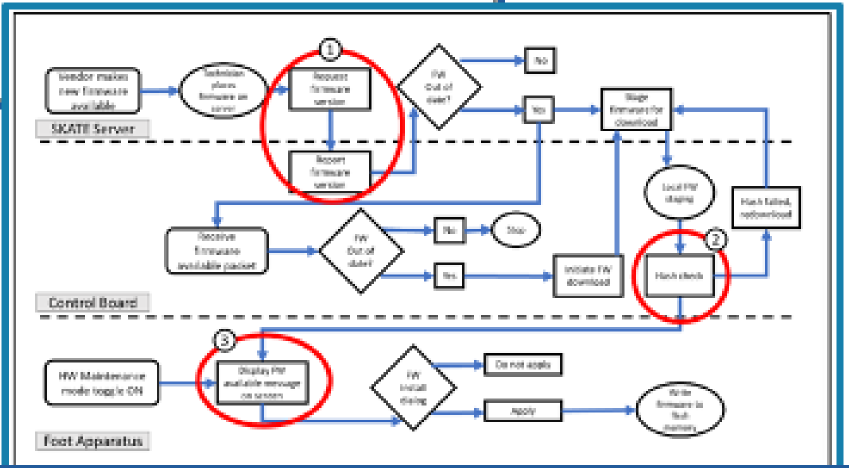
Diagraming High Value Dataflows


Client
Server
Authentication Database
Initial Connection
Username (U)+ Password(P) Request
Sends (U + P)
Forwards (U + P)
Provides Auth Response (R)
Provides Auth Response (R)

Swim Lane Diagram

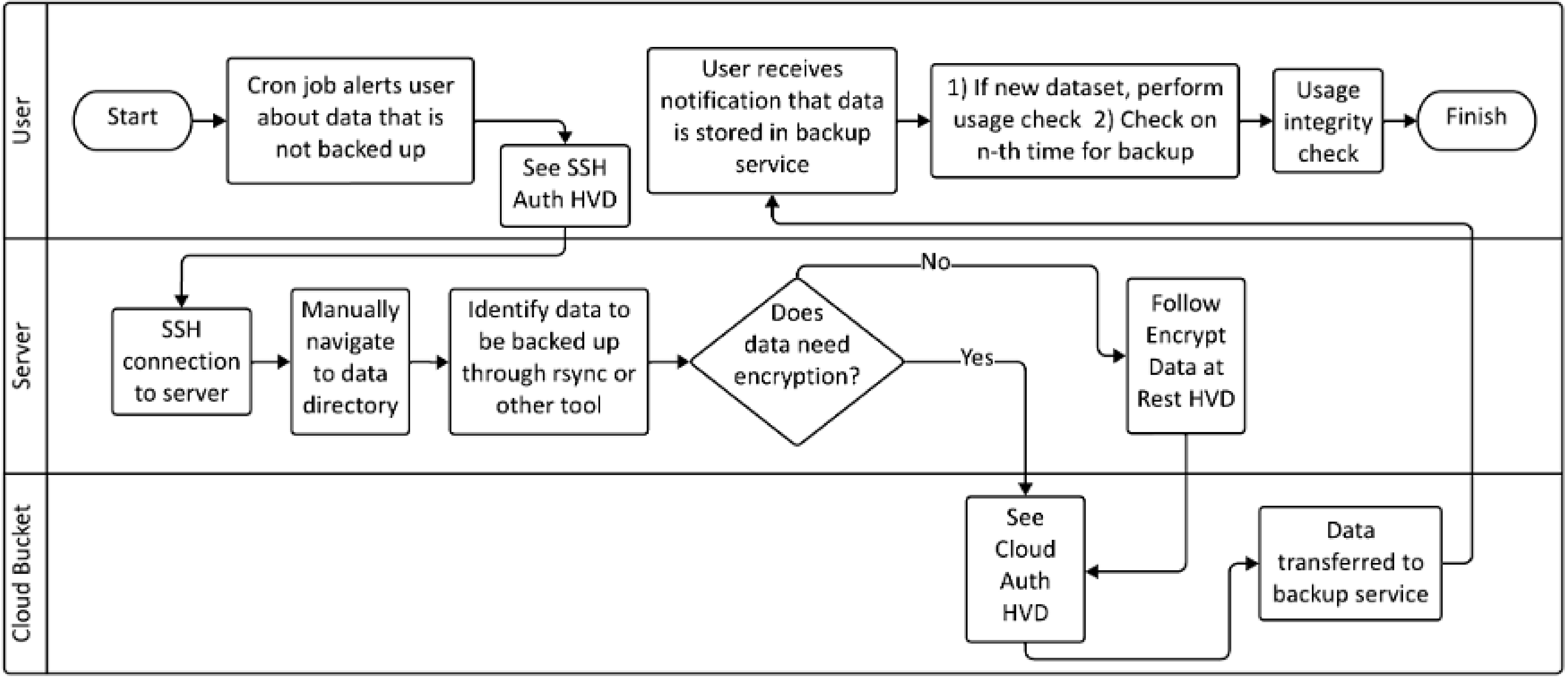
State Diagram


Failed Auth Sent
Server Terminates Connection
Client Sends Malformed Response
Credentials are Invalid
Server Awaiting Connection
Client Connection Initiated
Connection Created. Request For Auth Sent
Client Sends Well Formatted Credentials
Check Credentials For Database
Credentials are Valid
Session Established
No response Received



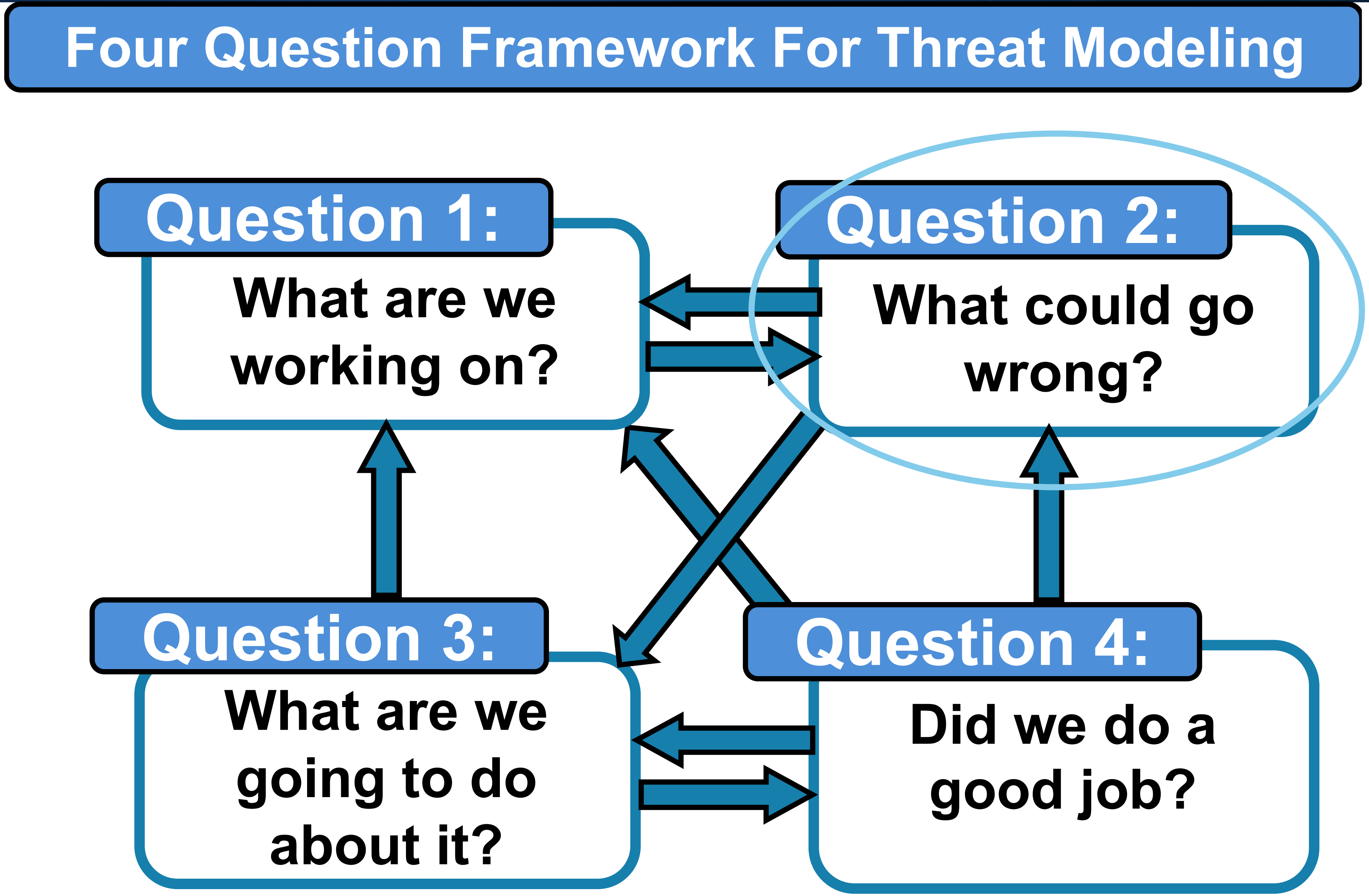
Cross Functional State Diagram

Cross-Functional Diagram Example: Backup Encrypted Data To Cloud Bucket


User
Server
Cloud Bucket
Start
Cron job alerts user about data that is not backed up
See SSH Auth HVD
User receives notification that data is stored in backup service
1) If new dataset, perform usage check 2) Check on n-th time for backup
Usage integrity check
Finish
SSH connection to server
Manually navigate to data directory
Identify data to be backed up through rsync or other tool
Does data need encryption?
No
Yes
Follow Encrypt Data at Rest HVD
See Cloud Auth HVD
Data transferred to backup service

Question 2

NIST | NATIONAL CYBERSECURITY CENTER OF EXCELLENCE


Four Question Framework For Threat Modeling

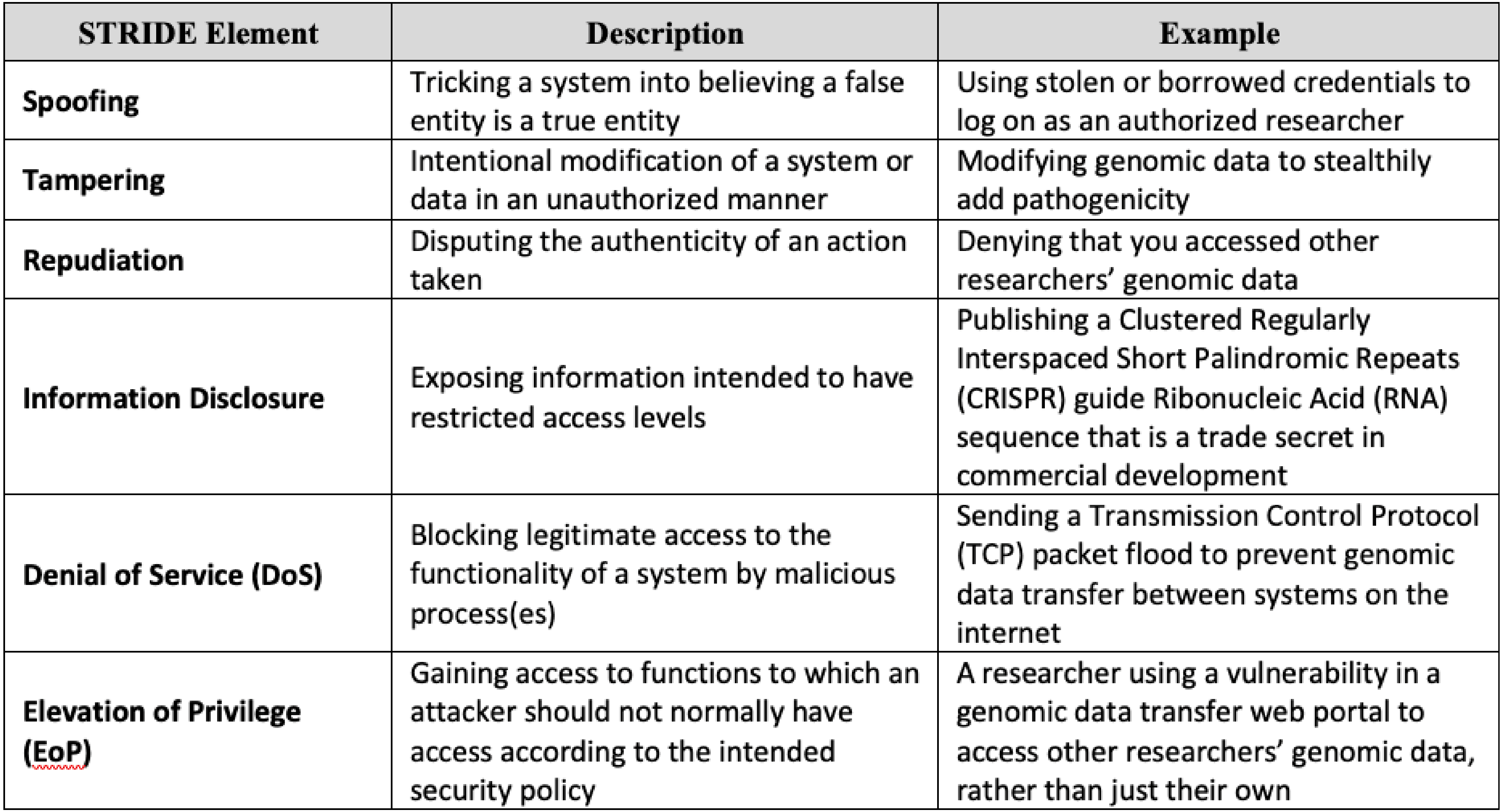
Question 1: What are we working on?

Question 2: What could go wrong?

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

Question 4: Did we do a good job?

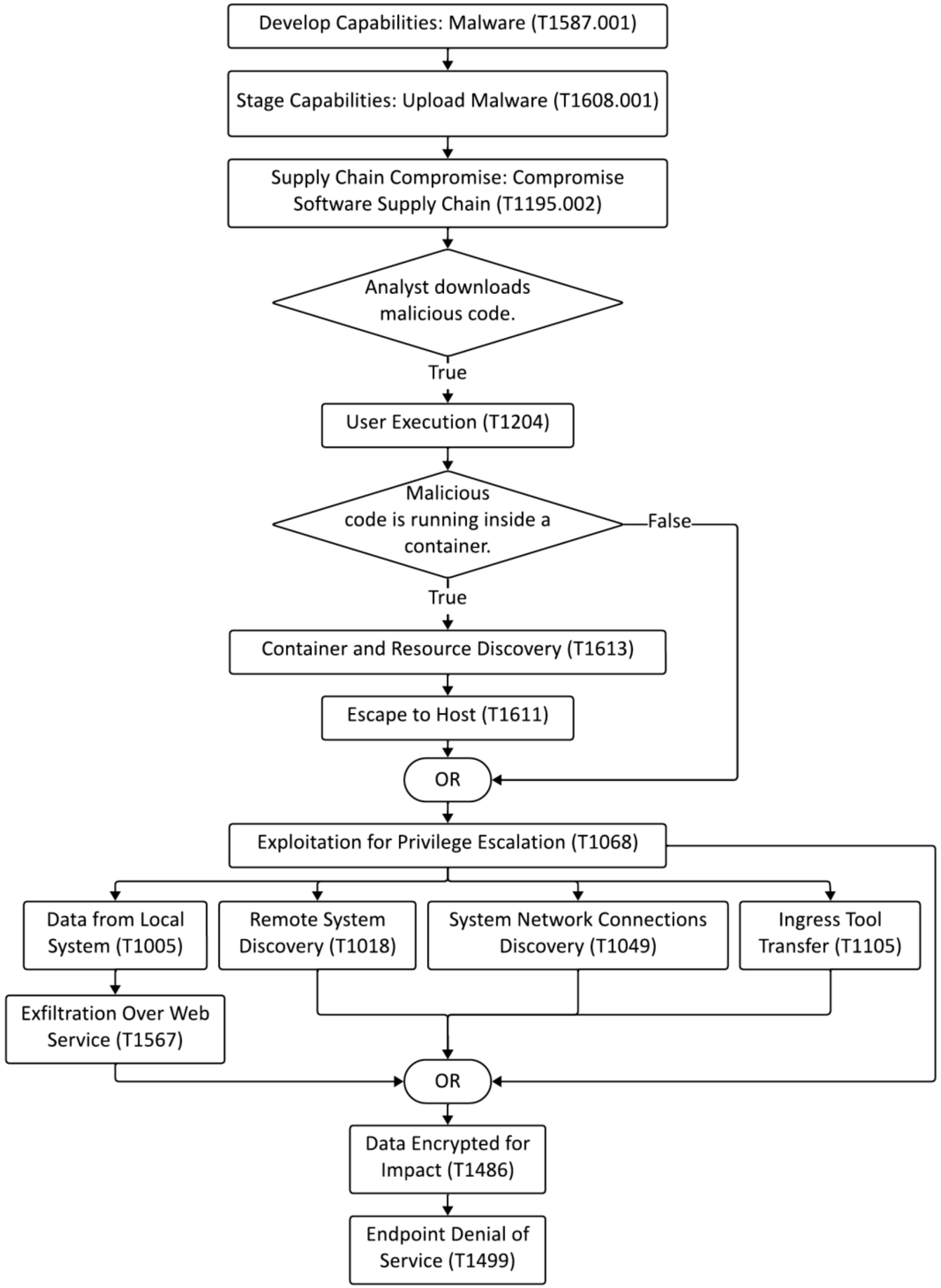
STRIDE (Description, Examples)



STRIDE Element	Description	Example
Spoofing	Tricking a system into believing a false entity is a true entity	Using stolen or borrowed credentials to log on as an authorized researcher
Tampering	Intentional modification of a system or data in an unauthorized manner	Modifying genomic data to stealthily add pathogenicity
Repudiation	Disputing the authenticity of an action taken	Denying that you accessed other researchers' genomic data
Information Disclosure	Exposing information intended to have restricted access levels	Publishing a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) guide Ribonucleic Acid (RNA) sequence that is a trade secret in commercial development
Denial of Service (DoS)	Blocking legitimate access to the functionality of a system by malicious process(es)	Sending a Transmission Control Protocol (TCP) packet flood to prevent genomic data transfer between systems on the internet
Elevation of Privilege (EoP)	Gaining access to functions to which an attacker should not normally have access according to the intended security policy	A researcher using a vulnerability in a genomic data transfer web portal to access other researchers' genomic data, rather than just their own

Untrusted Software Implanted with Malware

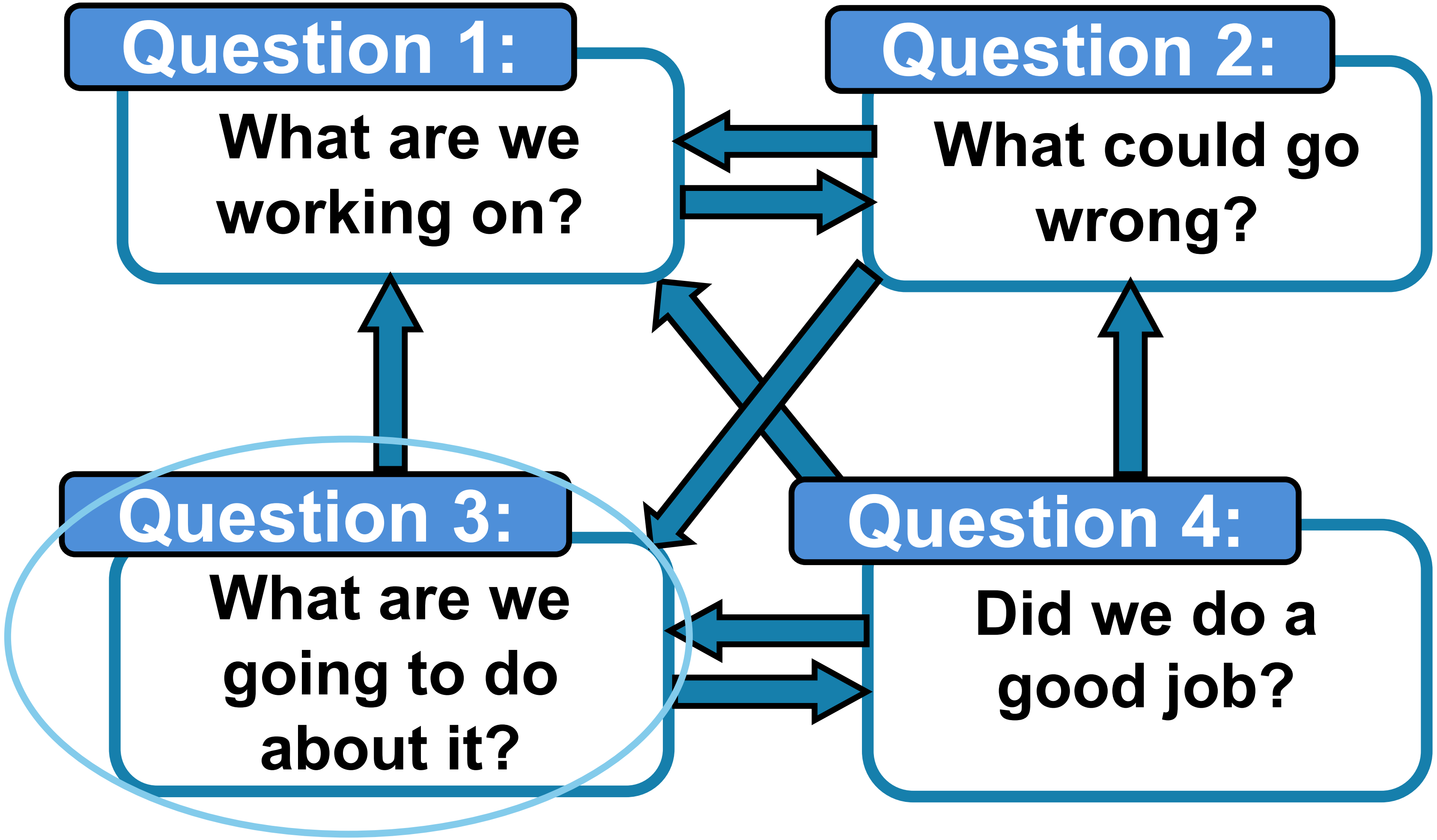
Attack Trees


Develop Capabilities: Malware (T1587.001)
Stage Capabilities: Upload Malware (T1608.001)
Supply Chain Compromise: Compromise Software Supply Chain (T1195.002)
Analyst downloads malicious code.
True
User Execution (T1204)
Malicious code is running inside a container.
False
True
Container and Resource Discovery (T1613)
Escape to Host (T1611)
OR
Exploitation for Privilege Escalation (T1068)
Data from Local System (T1005)
Remote System Discovery (T1018)
System Network Connections Discovery (T1049)
Ingress Tool Transfer (T1105)
Exfiltration Over Web Service (T1567)
OR
Data Encrypted for Impact (T1486)
Endpoint Denial of Service (T1499)

- Show how threats can be exploited
- Show that most cyber incidents require multiple steps
- Help in estimating likelihood and impact of an attack
- Can be analyzed to highlight locations for effective mitigations

Question 3

Four Question Framework For Threat Modeling



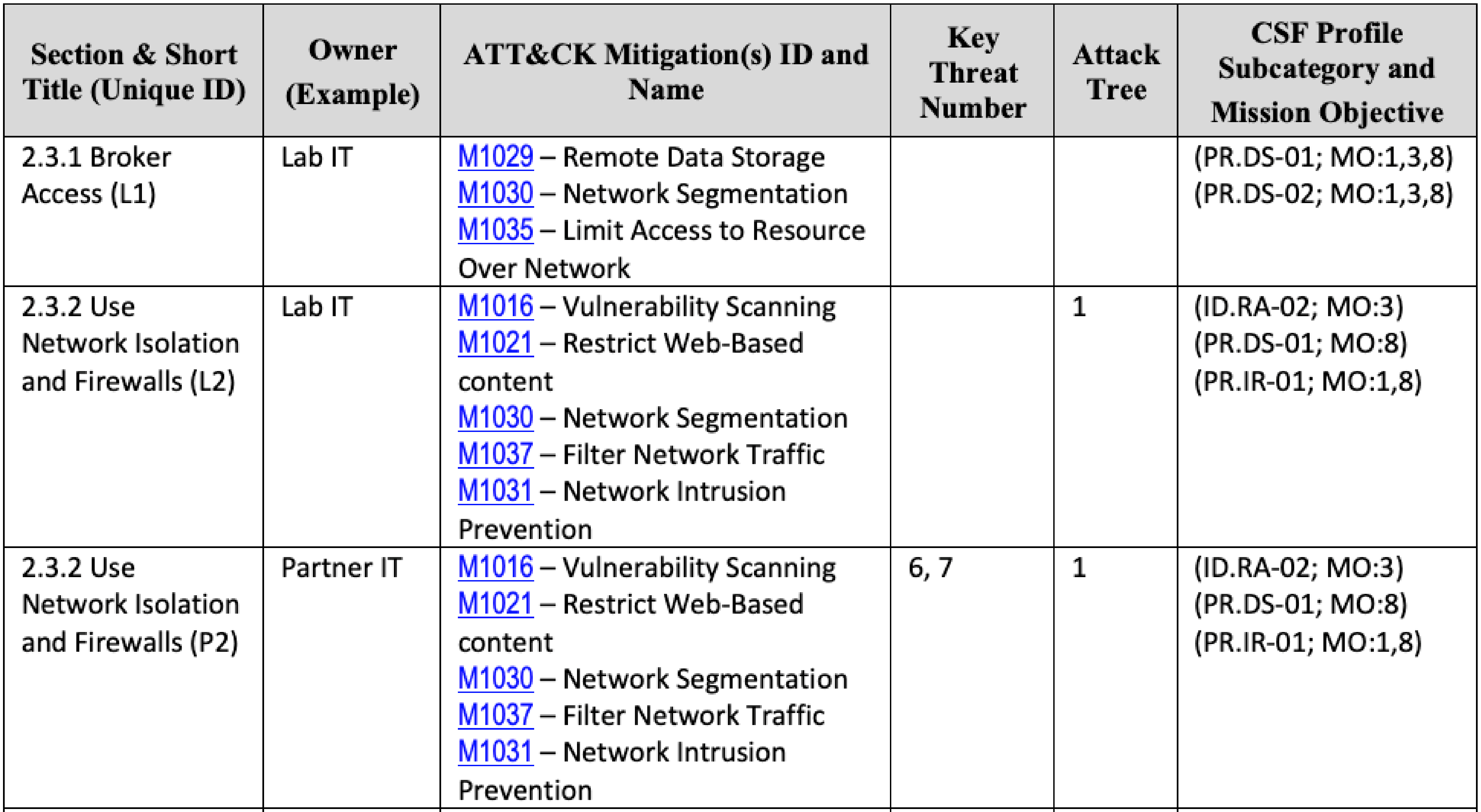
3. What are we going to do about it?

1. **Eliminate.** This is the most desired outcome; however, it is often challenging and may involve forgoing a specific feature or functionality. If that feature or function is required to accomplish the mission, then eliminating the threat is not possible.
2. **Mitigate.** This involves identifying, adding, and/or improving controls to protect, detect, respond to, or recover from attacks. For example, requiring multifactor authentication (MFA) would mitigate the threat of spoofing an authorized user.
3. **Accept.** In any system, there are unmitigated threats that cannot be eliminated or mitigated whose risk is judged to be acceptable. However, these accepted threats need to be documented and periodically reviewed.
4. **Transfer Responsibility.** This transfers the risk to another entity, who may have resources to mitigate (e.g., requiring users to use secure passwords or documenting the risk in an informed consent agreement) or who are willing to accept the risk.

Sample Mitigations

1. **Protect Genomic Data** (very large, valuable datasets)
   - Backup mitigations – tamper and denial of service threats
   - Encryption mitigation – exfiltration and tamper threats
   - Transfer mitigations – spoofing, elevation of privilege, repudiation, exfiltration, tamper threats
   - Role-Based Access Controls (RBAC) for genomic file systems
2. **Protect Sequencer** (very expensive equipment)
   - Micro segmentation
3. **Secure Analysis Software** (untrusted or open-source)
   - Containers with limited privileges for analysis software

Example Mitigation Table

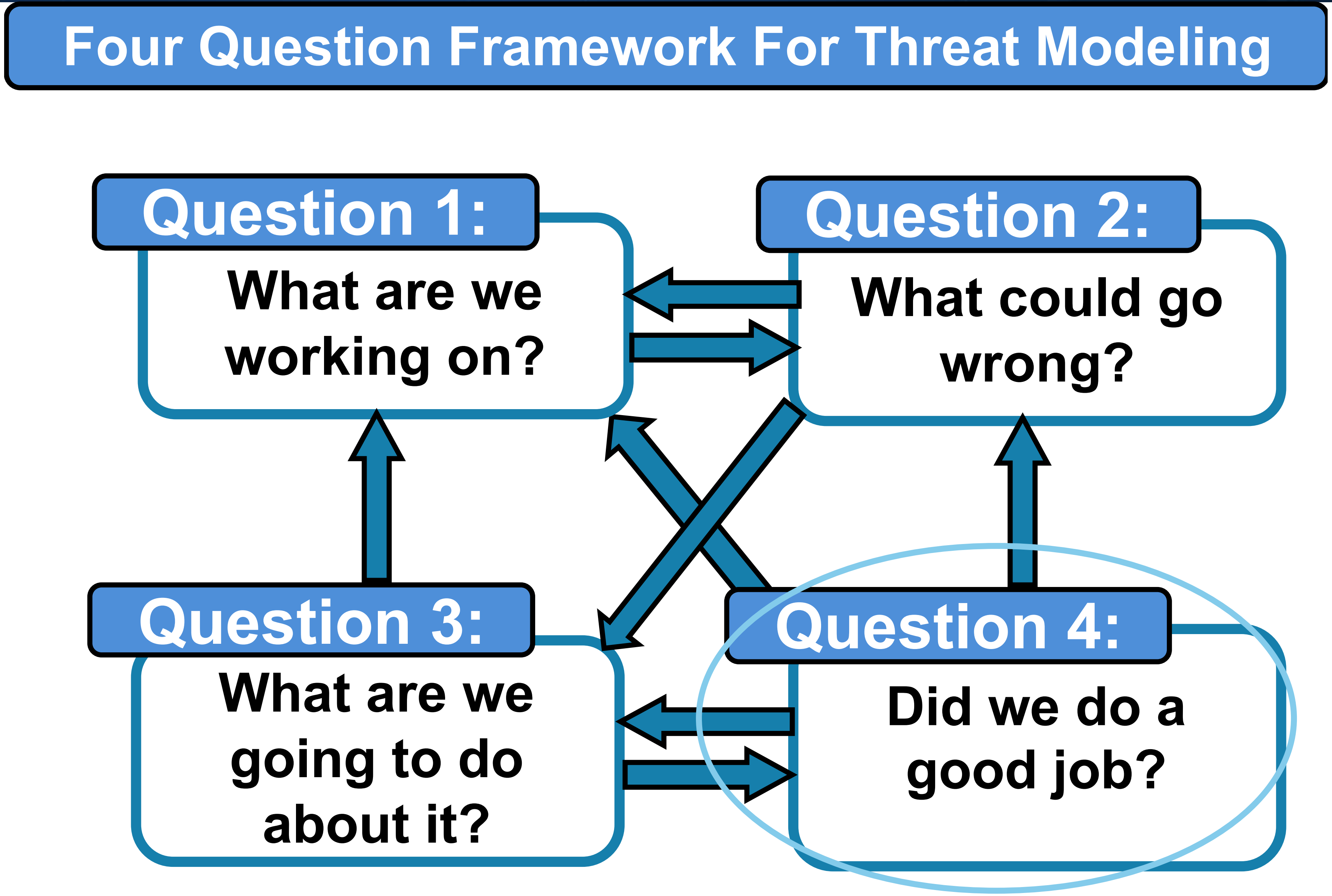


Section & Short Title (Unique ID)	Owner (Example)	ATT&CK Mitigation(s) ID and Name	Key Threat Number	Attack Tree	CSF Profile Subcategory and Mission Objective
2.3.1 Broker Access (L1)	Lab IT	M1029 – Remote Data Storage M1030 – Network Segmentation M1035 – Limit Access to Resource Over Network			(PR.DS-01; MO:1,3,8) (PR.DS-02; MO:1,3,8)
2.3.2 Use Network Isolation and Firewalls (L2)	Lab IT	M1016 – Vulnerability Scanning M1021 – Restrict Web-Based content M1030 – Network Segmentation M1037 – Filter Network Traffic M1031 – Network Intrusion Prevention		1	(ID.RA-02; MO:3) (PR.DS-01; MO:8) (PR.IR-01; MO:1,8)
2.3.2 Use Network Isolation and Firewalls (P2)	Partner IT	M1016 – Vulnerability Scanning M1021 – Restrict Web-Based content M1030 – Network Segmentation M1037 – Filter Network Traffic M1031 – Network Intrusion Prevention	6, 7	1	(ID.RA-02; MO:3) (PR.DS-01; MO:8) (PR.IR-01; MO:1,8)

Genome Sequencing Backup Options

Backup Option	File Size (GB) for 30x Human Genome on Illumina NovaSeq	Encryption Time	Notes on "Last Known Good State"	Cost per Year for Secondary Backup in AWS S3 Deep Glacier Flexible Retrieval
DNA in Freezer – sequence as needed	N/A	N/A	Pros – researcher might get to sequence on new tech Cons – cost is likely higher than data backup, not applicable for limited material	Freezer maintenance costs likely amortized over samples and other research projects
FASTQ.GZ (compressed reads file)	2 Files Reads 1 – 24GB Reads 2 – 25GB	Reads 1 – real 14m15.653s Reads 2 – real 16m15.559s	Requires re-mapping of reads that may be expensive if needed to perform for many samples	$2.12
BAM (mapped reads file)	37GB	real 22m30.975s user 1m11.776s sys 1m27.350s	May be the easiest to work from, but it locks a user into a reference genome and could require extra work	$1.60
CRAM (compressed mapped reads file)	14GB	real 8m8.965s user 0m25.891s sys 0m28.499s	Not as many analysis tools use a CRAM file as input compared to a BAM, so the user will need to know the impact on their pipeline	$0.60

Question 4



Four Question Framework For Threat Modeling
Question 1:
What are we working on?
Question 2:
What could go wrong?
Question 3:
What are we going to do about it?
Question 4:
Did we do a good job?

Did we do a good job on Q1 ?

Example questions for Question 1

- Is the DFD sufficiently detailed to capture communications between systems, particularly those that cross trust boundaries?
- Are all communications that cross trust boundaries included?
- Have HVDs been highlighted and is there sufficient model detail (in diagrams) to understand threats against them?
- Does the threat modeling explain how HVDs work and assess the impact of threats and mitigations to them?
- Are the diagram details sufficient or is additional information needed from suppliers, developers, or users?
- Are the DFD’s trust boundaries accurate and can they be enforced (for example, by network segmentation)?

Genomic Data Sequencing Unique Threats Identified

Initial threat modeling identified three areas where genomic data specific processing concerns and threats differed from most enterprise applications:

- Protections to address the unique ***value*** and potential ***size*** of genomic data, including closely managing ***remote access*** when sharing with external partners.
- Controls to protect the data when running ***untrusted researcher code*** during genomic data analysis.
- Safeguards to protect the highly valuable ***sequencers*** in the internal network that process sensitive genomic data and provide manufacturers access for maintenance.

Request for Comments and Looking Ahead

Feedback on NIST IR 8467

Welcome feedback and input on any aspect of NIST IR 2PD and additionally proposes a list of non-exhaustive questions and topics for consideration

1) How well do the practices in this publication relate to existing practices and standards leveraged by your organization? Are there significant gaps between the sets of practices that this publication should address? Would mappings to other standards and guidelines be useful or do mappings already exist that you use?
2) How do you expect this publication to influence your future practices and processes? Are there specific guidance documents or best practices that you recommend? For example, NIST published CSF 2.0 implementation examples to show potential ways to achieve the outcome in each Subcategory. What genomics-specific implementation examples would be valuable? What scenarios for cybersecurity and privacy threat modeling and mitigations would be valuable?
3) How do you envision using this publication? What changes would you like to see to increase/improve that use?
4) What suggestions do you have on changing the format of the information provided?
5) Is the guidance in this document sufficient to help your organization prioritize cybersecurity and privacy outcomes?

Feedback on CSWP 35

Welcome feedback and input on any aspect of NIST CSWP 35 and additionally proposes a list of non-exhaustive questions and topics for consideration

1) How well do the threat modeling practices in this white paper relate to existing threat modeling practices leveraged by your organization? Are there significant gaps between the sets of practices that this paper should address?
2) How do you expect this white paper to influence your future practices and processes?
3) How do you envision using this white paper? What changes would you like to see to increase/improve that use?
4) What suggestions do you have on changing the format of the information provided? Would it help to provide a more concise overview document with additional detail provided in either appendices or as part of a more interactive website (e.g., GitHub Pages as used in the NCCoE Zero Trust project)?
5) Is the example provided here sufficient for your organization to identify and address cybersecurity threats in genomic data sequencing or genomic data analysis? Are there changes or additional content that the authors should consider?

Looking ahead

- Provide input on the document through the Public Comments by January 30, 2025
- Consider attending workshop (in-person or virtual) on May 20th at NCCoE in Rockville, MD

Cybersecurity of Genomic Data Project Page

https://www.nccoe.nist.gov/projects/cybersecurity-genomic-data

Fact Sheet: https://www.nccoe.nist.gov/sites/default/files/2024-12/genomics-fact-sheet.pdf

Email: genomic_cybersecurity_nccoe@nist.gov

PIs: Ronald Pulivarti (ronald.pulivarti@nist.gov)

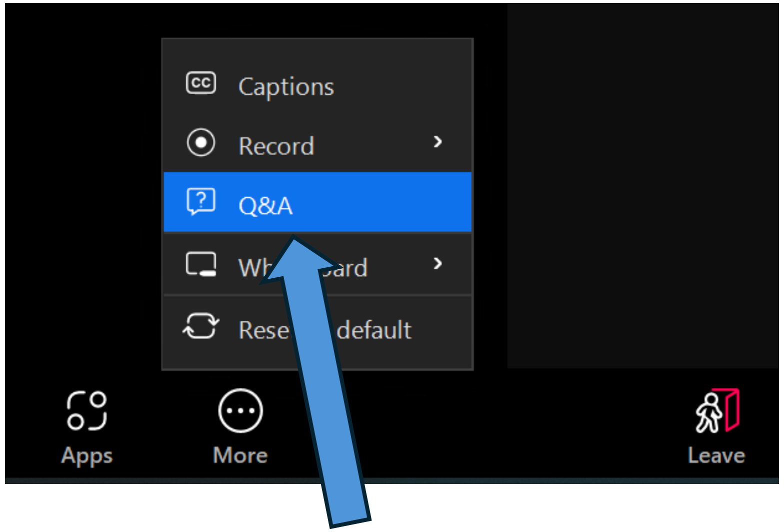
Justin Wagner (justin.wagner@nist.gov)

Q&A

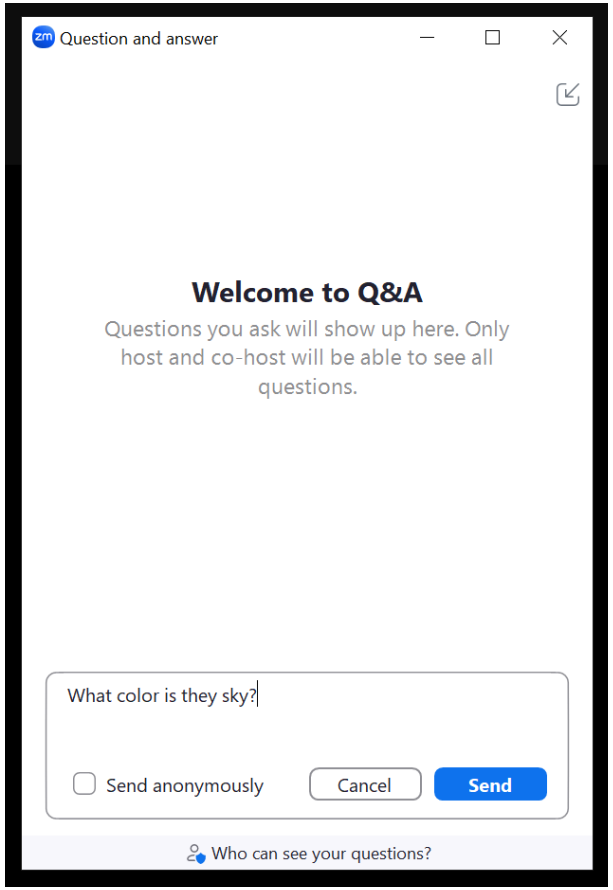
Submitting Questions

Please use the Q&A function to enter your questions.

We will do our best to answer all questions during the Q&A session at the end of this event.



1. To open the Q&A function, click on “More” at the bottom of your screen and select the “Q&A” option.



Welcome to Q&A

Questions you ask will show up here. Only host and co-host will be able to see all questions.

What color is they sky?

2. Type your question in the text box and click Send

Backup

Did we do a good job on Q2?

Actions that could improve threat identification (Question 2)

- Review organizational policies, strategies, and processes to determine if there are other threat areas not being addressed by the technical evaluation.
- Address missing STRIDE Elements.
- Consider attacks that have occurred in the genomic stakeholder community and closely adjacent industries. Threat intelligence can be used to identify TTPs favored by actors who are known to target an industry.
- Consider known vulnerabilities in software and services being used.
- Determine whether the threats being considered map to the threats listed in published documents for the genomic community, such as ***NIST IR 8432.***

Did we do a good job on Q3?

Example questions and actions to evaluate Question 3

- How thorough are the mitigations? Regularly consider the impact of any changes to the system or threat environment.
- Is there a risk strategy for every threat that crosses a trust boundary?
- Consider additional mitigations and other responses across each risk strategy: eliminate, accept, and transfer.
- Conduct a legal review to determine if the mitigations, accepted risks, and transferred risks (particularly the manner of transfer notification) meet the necessary regulatory requirements.
- Perform a security assessment including automated and manual penetration testing to evaluate how the mitigations and threat modeling perform and how they can be improved.