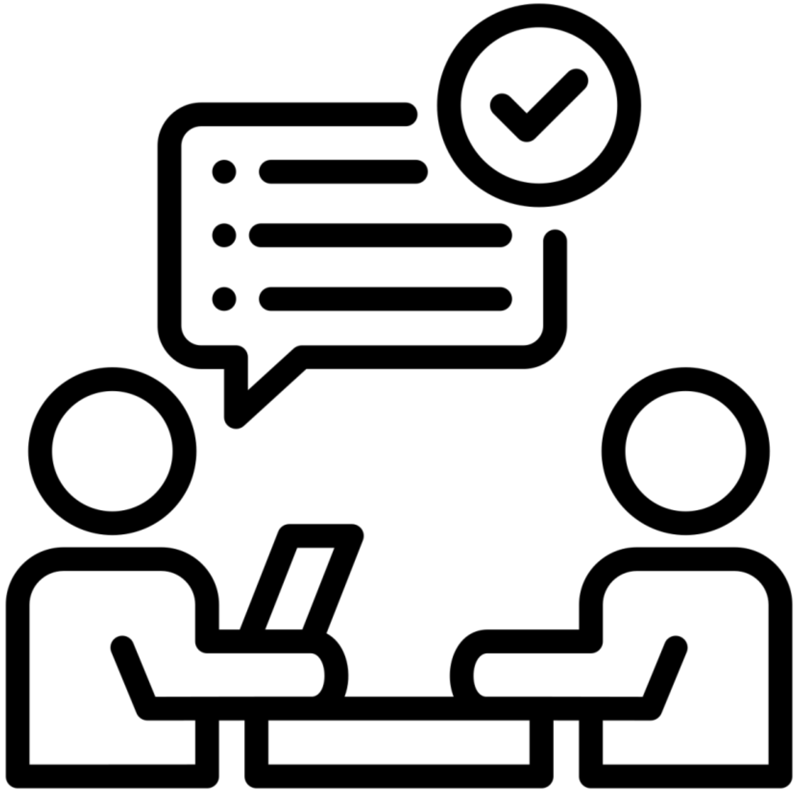
Privacy Threat Modeling in Practice

Cat Easdon, Privacy Engineering Team Lead, Dynatrace

Why threat model?

Impact assessments

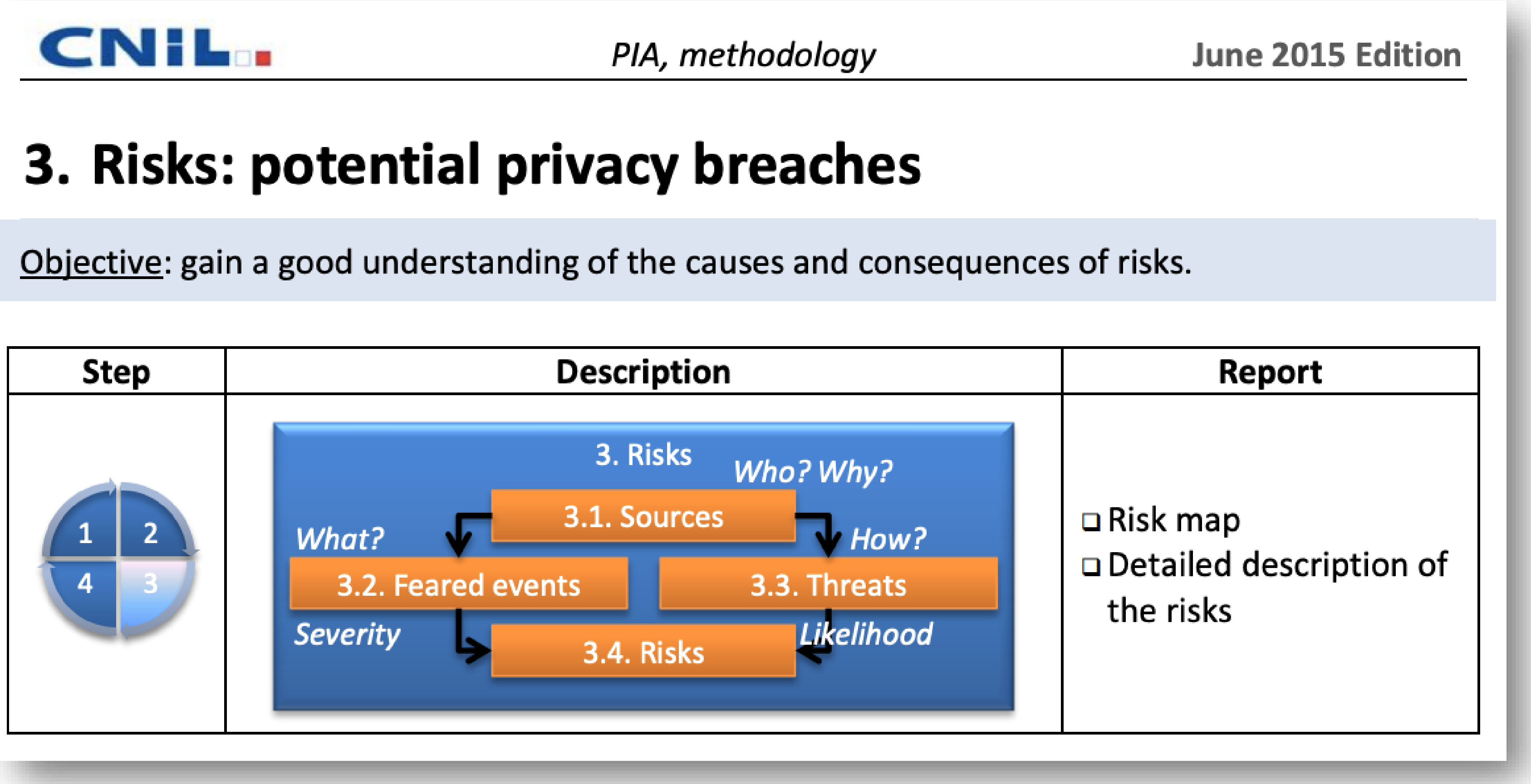
- GDPR: **data protection impact assessment** (DPIA) required if a new processing activity is likely to result in high risk
  - Also known as a **privacy impact assessment** (PIA)
- Common in risk-based digital regulation, e.g. the AI Act’s **fundamental rights impact assessment** (FRIA)



High-level process overview

1. Determine that an impact assessment is needed
2. Document the data flows and processing activities
    1. Purpose for processing, legal basis, necessity, proportionality, etc.
3. **Identify and evaluate risks to natural persons** *(Today's focus)*
4. Identify existing mitigations
5. Identify and prioritize additional necessary mitigations
6. Sign-off (with action plan)

Risk assessment



CNIL *PIA, methodology* June 2015 Edition

3. Risks: potential privacy breaches

Objective: gain a good understanding of the causes and consequences of risks.

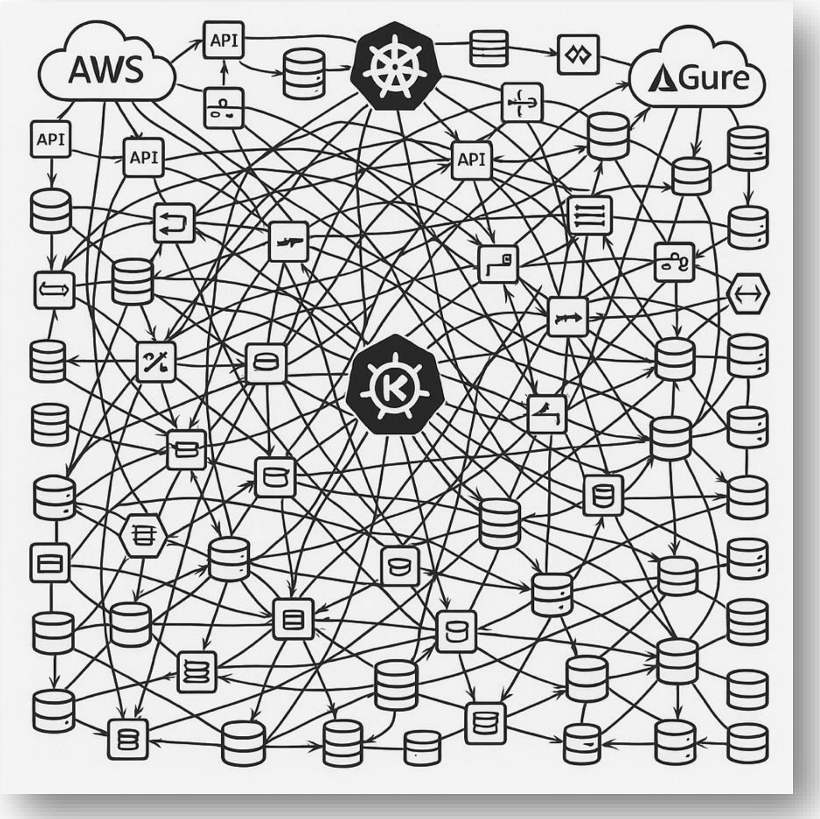
Step	Description	Report
1 2 4 3	3. Risks Who? Why? 3.1. Sources What? How? 3.2. Feared events 3.3. Threats Severity Likelihood 3.4. Risks	❑ Risk map ❑ Detailed description of the risks

Risk assessment is hard!

Dariusz Kloza, Thibaut D'hulst and Malik Aouadi, [What could possibly go wrong? On risks to the rights and freedoms of natural persons in EU data protection law, their typologies and their identification](), Technology and Regulation, 2024

> "…given the complexity of contemporary processing operations, the **elusive nature of risk** often makes it difficult for data subjects and for controllers and processors alike to have a clear and comprehensive understanding of **what could possibly go wrong**."

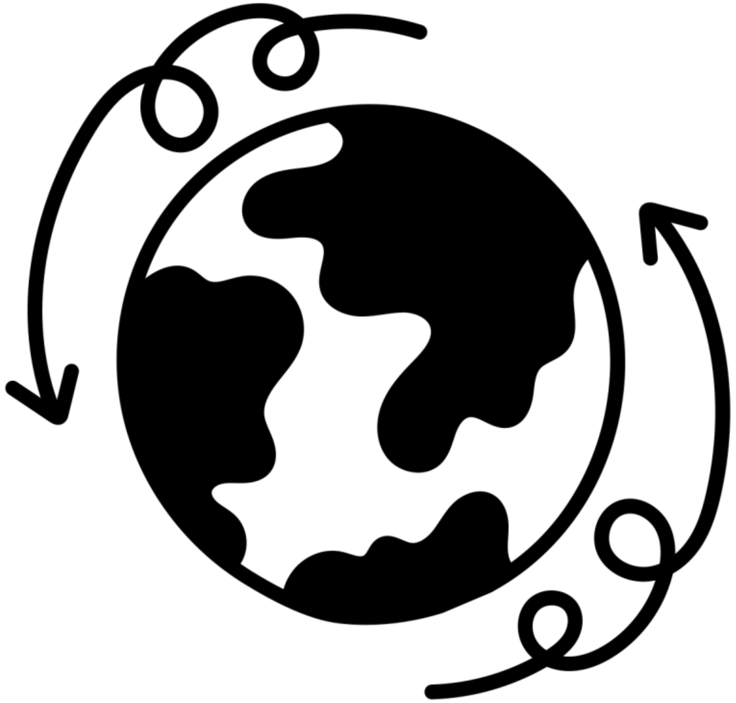
Risk assessment is hard!



A dataflow diagram that even an AI can't comprehend

Our motivating example today

- Your company is acquiring a company that provides encrypted email services
- Integrating their product with your SSO will mean that timestamped EU customer IP addresses are transferred to **Boggleland** on each login
- You are conducting a DPIA and are unsure how to determine:
  - What could go wrong for data subjects?
  - How bad would it be (for us and them!)?
  - Should we do anything about it?



Enter threat modeling...

THREAT MODELING MANIFESTO

Those questions sound like the threat modeling manifesto:

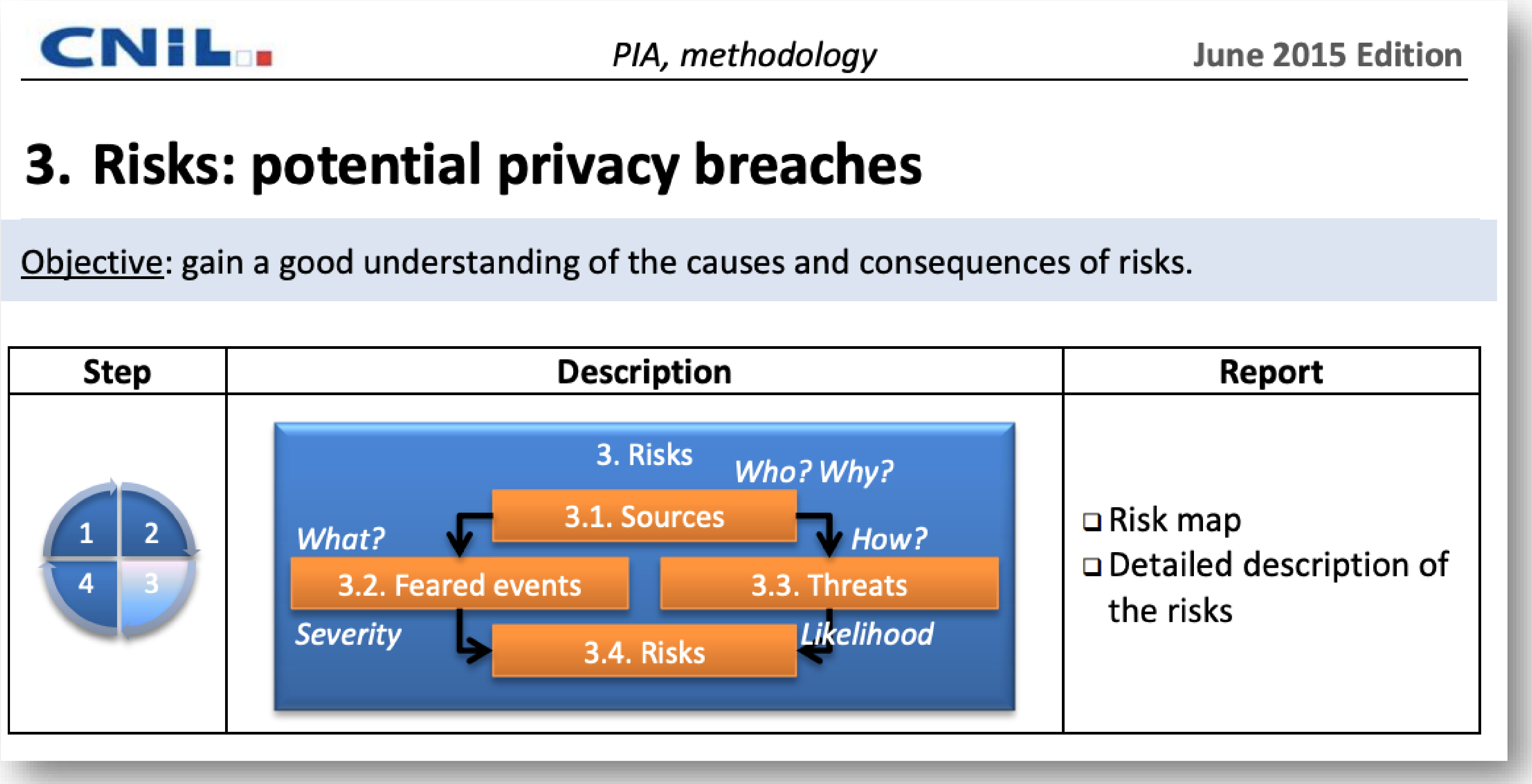
1. What are we working on?
2. What can go wrong?
3. What are we going to do about it?
4. Did we do a good enough job?

Core threat modeling concepts

THREAT MODELING MANIFESTO

- Anyone can threat model!
- No one system diagram or single person’s understanding will be perfect – create multiple diagrams and collaborate with multiple stakeholders
- Conduct early and frequent analysis; iterate and improve over time

Risk assessment revisited



CNIL PIA, methodology June 2015 Edition

3. Risks: potential privacy breaches

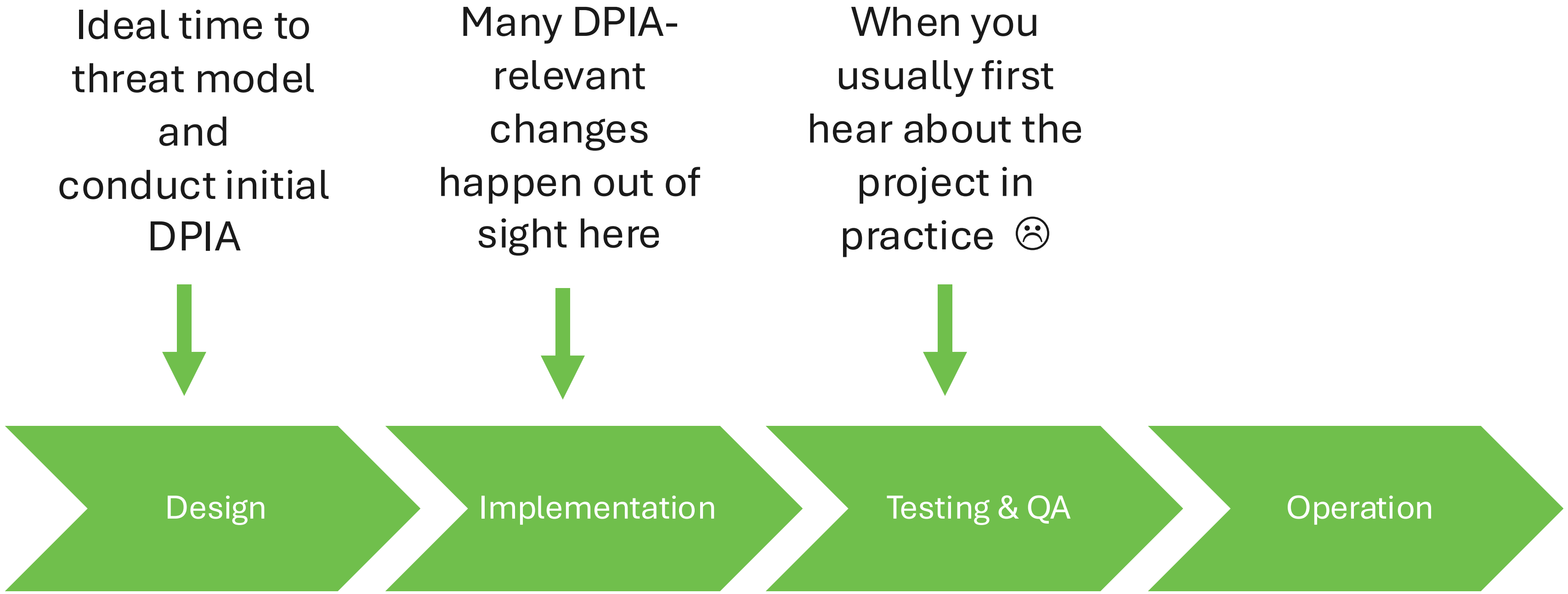
Objective: gain a good understanding of the causes and consequences of risks.

Step	Description	Report
1 2 4 3	3. Risks Who? Why? 3.1. Sources What? How? 3.2. Feared events 3.3. Threats Severity Likelihood 3.4. Risks	❑ Risk map ❑ Detailed description of the risks

Risk assessment with threat modeling

- **Risk identification** (*who*, *why*, *what*, and *how*): collaborative threat modeling
- **Risk evaluation:** risk = *severity/impact * likelihood*
  - Simple, right? Just calculate and prioritize accordingly 😜
  - Unfortunately very context-specific – more on this later

Threat modeling in the software development lifecycle



Ideal time to threat model and conduct initial DPIA

Many DPIA-relevant changes happen out of sight here

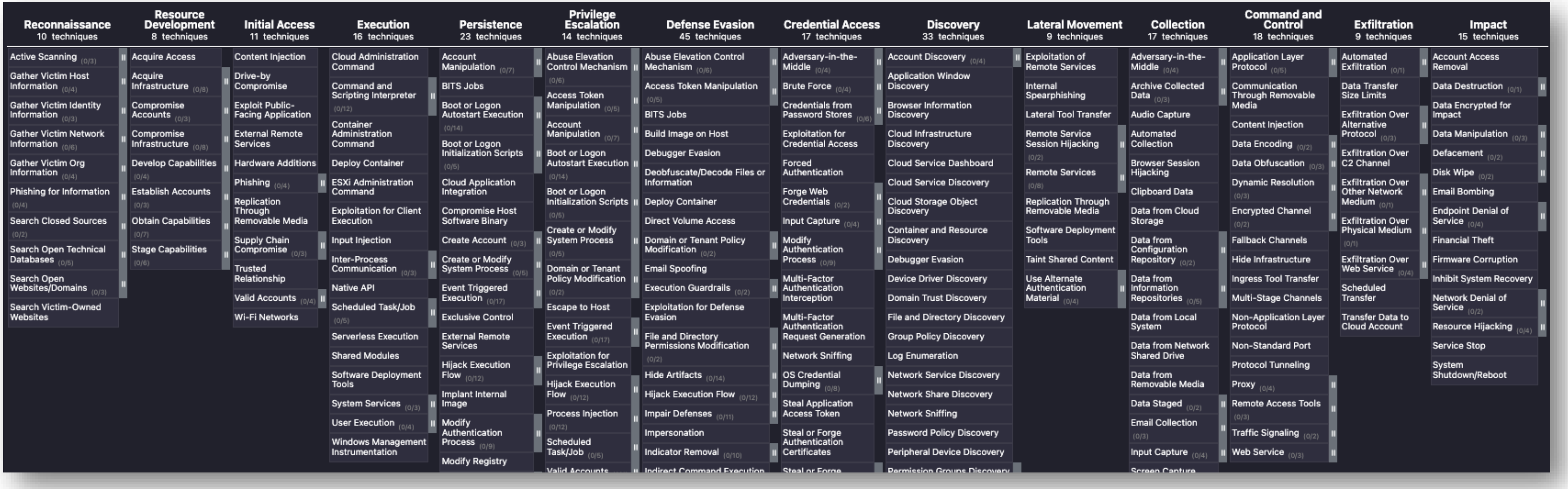
When you usually first hear about the project in practice ☹

Threat modeling in security

STRIDE

- Spoofing
- Tampering
- Repudiation
- Information Disclosure
- Denial of Service
- Elevation of Privilege

MITRE ATT&CK®



Reconnaissance 10 techniques	Resource Development 8 techniques	Initial Access 11 techniques	Execution 16 techniques	Persistence 23 techniques	Privilege Escalation 14 techniques	Defense Evasion 45 techniques	Credential Access 17 techniques	Discovery 33 techniques	Lateral Movement 9 techniques	Collection 17 techniques	Command and Control 18 techniques	Exfiltration 9 techniques	Impact 15 techniques
Active Scanning (0/3)	Acquire Access	Content Injection	Cloud Administration Command	Account Manipulation (0/7)	Abuse Elevation Control Mechanism (0/6)	Abuse Elevation Control Mechanism (0/6)	Adversary-in-the-Middle (0/4)	Account Discovery (0/4)	Exploitation of Remote Services	Adversary-in-the-Middle (0/4)	Application Layer Protocol (0/5)	Automated Exfiltration (0/1)	Account Access Removal
Gather Victim Host Information (0/4)	Acquire Infrastructure (0/8)	Drive-by Compromise	Command and Scripting Interpreter (0/12)	BITS Jobs	Access Token Manipulation (0/5)	Access Token Manipulation (0/5)	Brute Force (0/4)	Application Window Discovery	Internal Spearphishing	Archive Collected Data (0/3)	Communication Through Removable Media	Data Transfer Size Limits	Data Destruction (0/1)
Gather Victim Identity Information (0/3)	Compromise Accounts (0/3)	Exploit Public-Facing Application	Container Administration Command	Boot or Logon Autostart Execution (0/14)	Account Manipulation (0/7)	BITS Jobs	Credentials from Password Stores (0/6)	Browser Information Discovery	Lateral Tool Transfer	Audio Capture	Content Injection	Exfiltration Over Alternative Protocol (0/3)	Data Encrypted for Impact
Gather Victim Network Information (0/6)	Compromise Infrastructure (0/8)	External Remote Services	Deploy Container	Boot or Logon Initialization Scripts (0/5)	Boot or Logon Autostart Execution (0/14)	Build Image on Host	Exploitation for Credential Access	Cloud Infrastructure Discovery	Remote Service Session Hijacking (0/2)	Automated Collection	Data Encoding (0/2)	Exfiltration Over C2 Channel	Data Manipulation (0/3)
Gather Victim Org Information (0/4)	Develop Capabilities (0/4)	Hardware Additions	ESXi Administration Command	Cloud Application Integration	Boot or Logon Initialization Scripts (0/5)	Debugger Evasion	Forced Authentication	Cloud Service Dashboard	Remote Services (0/8)	Browser Session Hijacking	Data Obfuscation (0/3)	Exfiltration Over Other Network Medium (0/1)	Defacement (0/2)
Phishing for Information (0/4)	Establish Accounts (0/3)	Phishing (0/4)	Exploitation for Client Execution	Compromise Host Software Binary	Create or Modify System Process (0/5)	Deobfuscate/Decode Files or Information	Forge Web Credentials (0/2)	Cloud Service Discovery	Replication Through Removable Media	Clipboard Data	Dynamic Resolution (0/3)	Exfiltration Over Physical Medium (0/1)	Disk Wipe (0/2)
Search Closed Sources (0/2)	Obtain Capabilities (0/7)	Replication Through Removable Media	Input Injection	Create Account (0/3)	Domain or Tenant Policy Modification (0/2)	Deploy Container	Input Capture (0/4)	Cloud Storage Object Discovery	Software Deployment Tools	Data from Cloud Storage	Encrypted Channel (0/2)	Exfiltration Over Web Service (0/4)	Email Bombing
Search Open Technical Databases (0/5)	Stage Capabilities (0/6)	Supply Chain Compromise (0/3)	Inter-Process Communication (0/3)	Create or Modify System Process (0/5)	Escape to Host	Direct Volume Access	Modify Authentication Process (0/9)	Container and Resource Discovery	Taint Shared Content	Data from Configuration Repository (0/2)	Fallback Channels	Scheduled Transfer	Endpoint Denial of Service (0/4)
Search Open Websites/Domains (0/3)		Trusted Relationship	Native API	Event Triggered Execution (0/17)	Event Triggered Execution (0/17)	Domain or Tenant Policy Modification (0/2)	Multi-Factor Authentication Interception	Debugger Evasion	Use Alternate Authentication Material (0/4)	Data from Information Repositories (0/5)	Hide Infrastructure	Transfer Data to Cloud Account	Financial Theft
Search Victim-Owned Websites		Valid Accounts (0/4)	Scheduled Task/Job (0/5)	Exclusive Control	Exploitation for Privilege Escalation	Email Spoofing	Multi-Factor Authentication Request Generation	Device Driver Discovery		Data from Local System	Ingress Tool Transfer		Firmware Corruption
		Wi-Fi Networks	Serverless Execution	External Remote Services	Hijack Execution Flow (0/12)	Execution Guardrails (0/2)	Network Sniffing	Domain Trust Discovery		Data from Network Shared Drive	Multi-Stage Channels		Inhibit System Recovery
			Shared Modules	Hijack Execution Flow (0/12)	Process Injection (0/12)	Exploitation for Defense Evasion	OS Credential Dumping (0/8)	File and Directory Discovery		Data from Removable Media	Non-Application Layer Protocol		Network Denial of Service (0/2)
			Software Deployment Tools	Implant Internal Image	Scheduled Task/Job (0/5)	File and Directory Permissions Modification (0/2)	Steal Application Access Token	Group Policy Discovery		Data Staged (0/2)	Non-Standard Port		Resource Hijacking (0/4)
			System Services (0/3)	Modify Authentication Process (0/9)	Valid Accounts	Hide Artifacts (0/14)	Steal or Forge Authentication Certificates	Log Enumeration		Email Collection (0/3)	Protocol Tunneling		Service Stop
			User Execution (0/4)	Modify Registry		Hijack Execution Flow (0/12)	Steal or Forge	Network Service Discovery		Input Capture (0/4)	Proxy (0/4)		System Shutdown/Reboot
			Windows Management Instrumentation			Impair Defenses (0/11)		Network Share Discovery		Screen Capture	Remote Access Tools (0/3)		
						Impersonation		Network Sniffing			Traffic Signaling (0/2)		
						Indicator Removal (0/10)		Password Policy Discovery			Web Service (0/3)		
						Indirect Command Execution		Peripheral Device Discovery					
								Permission Groups Discovery					

Privacy threat modeling: LINDDUN

LINDDUN

- Linking
- Identifying
- Non-Repudiation
- Non-Compliance
- Data Disclosure
- Detecting
- Unawareness & Unintervenability

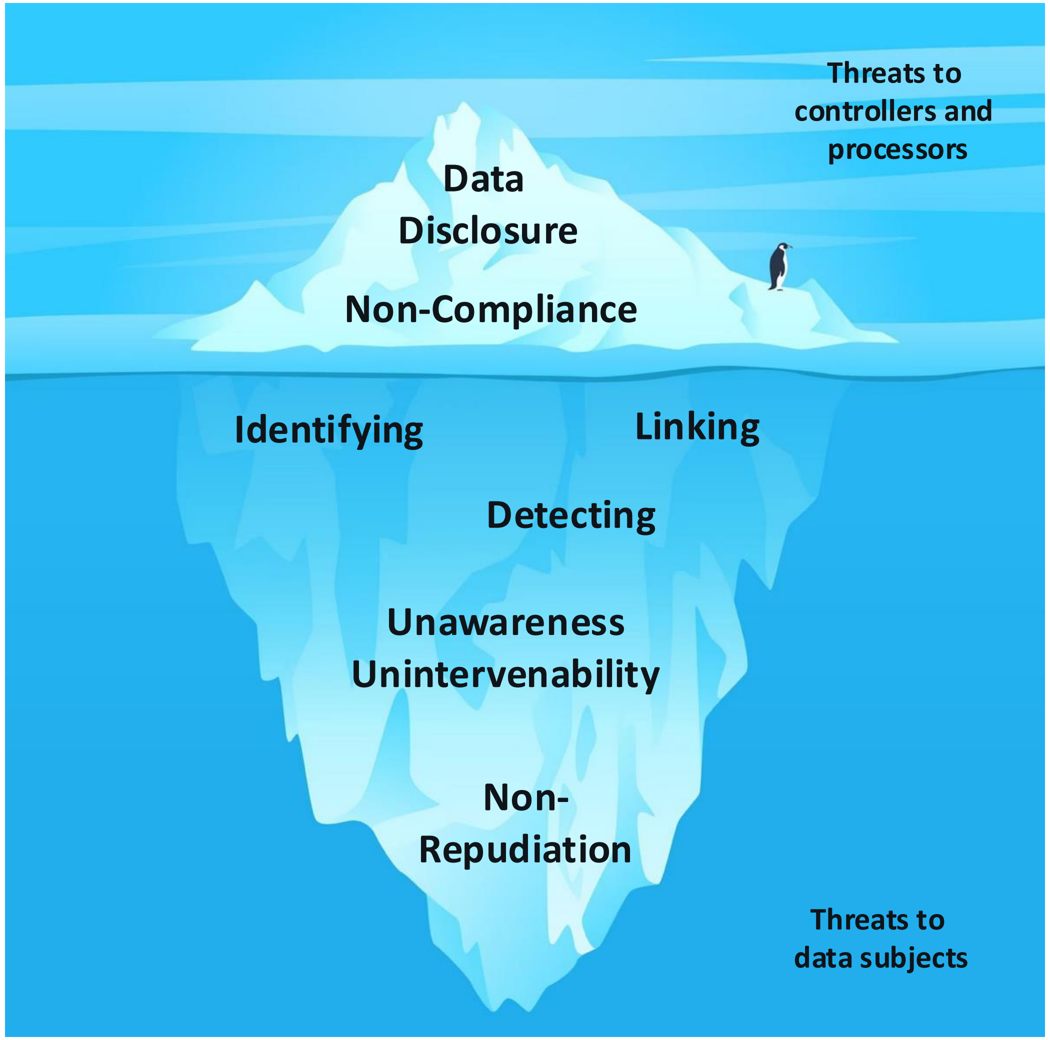
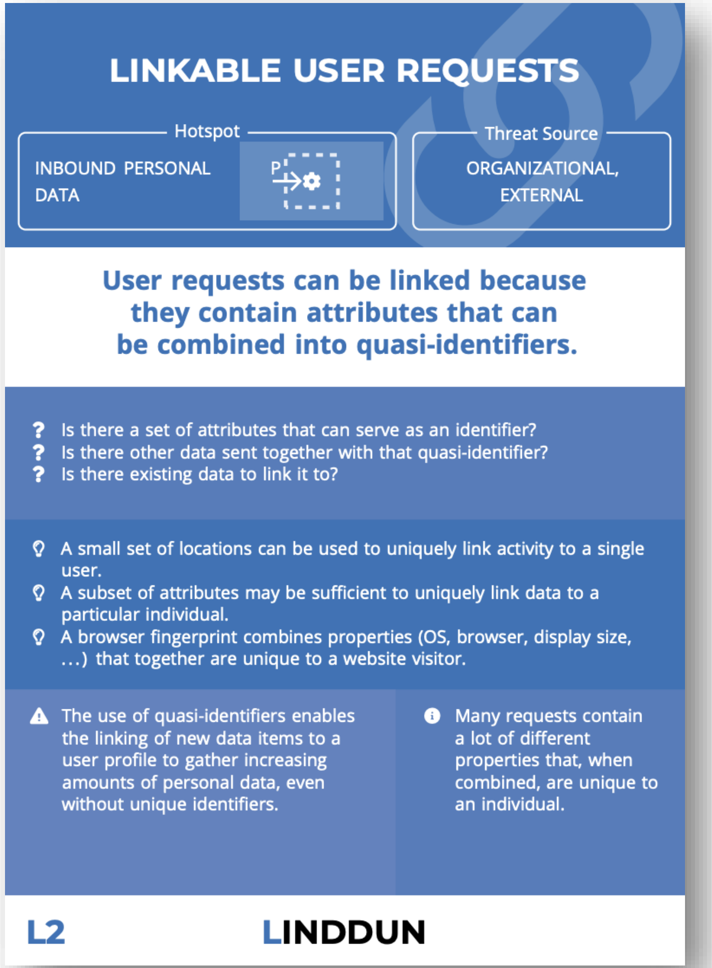


Image designed by Freepik



LINKABLE USER REQUESTS

Hotspot: INBOUND PERSONAL DATA

Threat Source: ORGANIZATIONAL, EXTERNAL

User requests can be linked because they contain attributes that can be combined into quasi-identifiers.

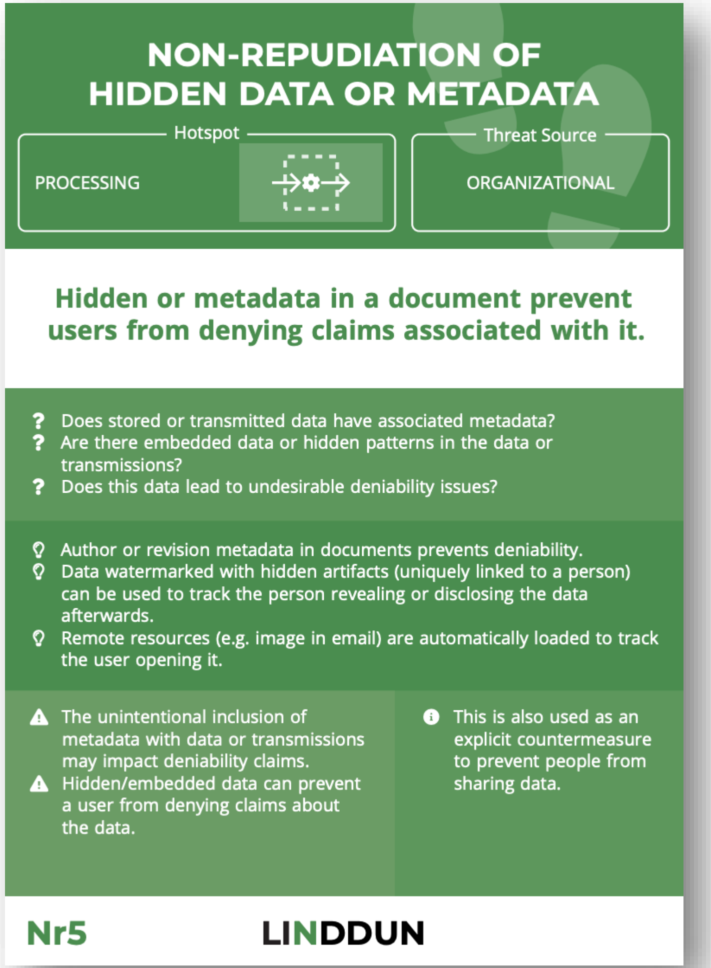
- ? Is there a set of attributes that can serve as an identifier?
- ? Is there other data sent together with that quasi-identifier?
- ? Is there existing data to link it to?

- 💡 A small set of locations can be used to uniquely link activity to a single user.
- 💡 A subset of attributes may be sufficient to uniquely link data to a particular individual.
- 💡 A browser fingerprint combines properties (OS, browser, display size, …) that together are unique to a website visitor.

- ⚠ The use of quasi-identifiers enables the linking of new data items to a user profile to gather increasing amounts of personal data, even without unique identifiers.

- ℹ Many requests contain a lot of different properties that, when combined, are unique to an individual.

L2 LINDDUN



NON-REPUDIATION OF HIDDEN DATA OR METADATA

Hotspot: PROCESSING

Threat Source: ORGANIZATIONAL

Hidden or metadata in a document prevent users from denying claims associated with it.

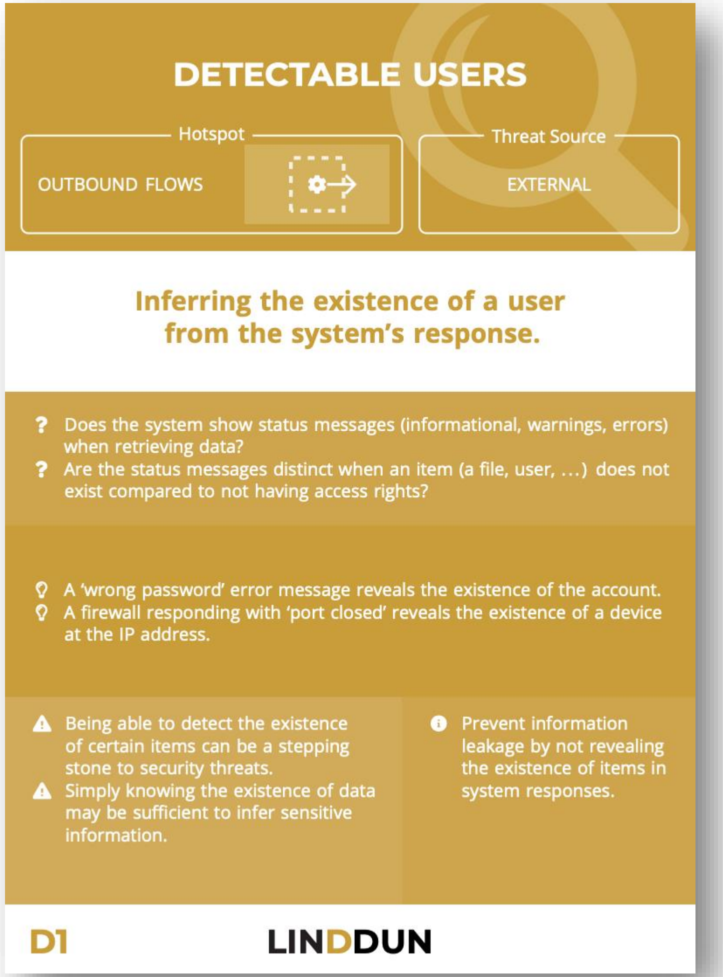
- ? Does stored or transmitted data have associated metadata?
- ? Are there embedded data or hidden patterns in the data or transmissions?
- ? Does this data lead to undesirable deniability issues?

- 💡 Author or revision metadata in documents prevents deniability.
- 💡 Data watermarked with hidden artifacts (uniquely linked to a person) can be used to track the person revealing or disclosing the data afterwards.
- 💡 Remote resources (e.g. image in email) are automatically loaded to track the user opening it.

- ⚠ The unintentional inclusion of metadata with data or transmissions may impact deniability claims.
- ⚠ Hidden/embedded data can prevent a user from denying claims about the data.

- ℹ This is also used as an explicit countermeasure to prevent people from sharing data.

Nr5 LINDDUN



DETECTABLE USERS

Hotspot: OUTBOUND FLOWS

Threat Source: EXTERNAL

Inferring the existence of a user from the system's response.

- ? Does the system show status messages (informational, warnings, errors) when retrieving data?
- ? Are the status messages distinct when an item (a file, user, …) does not exist compared to not having access rights?

- 💡 A 'wrong password' error message reveals the existence of the account.
- 💡 A firewall responding with 'port closed' reveals the existence of a device at the IP address.

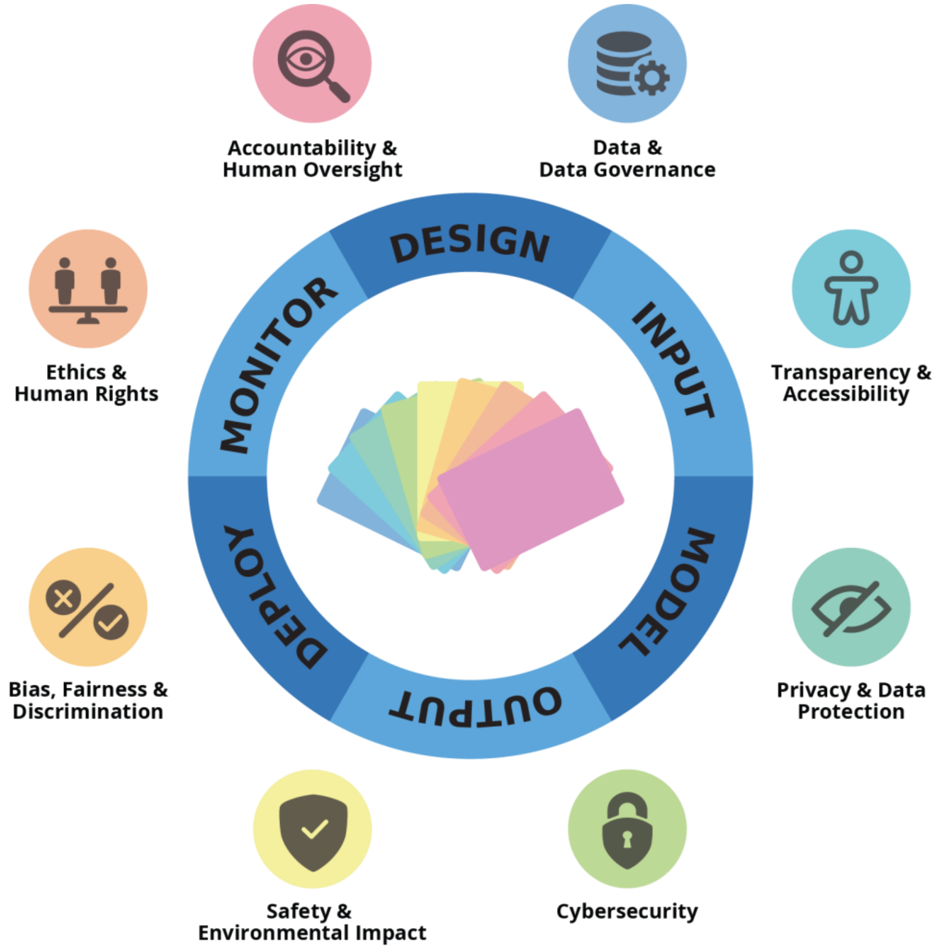
- ⚠ Being able to detect the existence of certain items can be a stepping stone to security threats.
- ⚠ Simply knowing the existence of data may be sufficient to infer sensitive information.

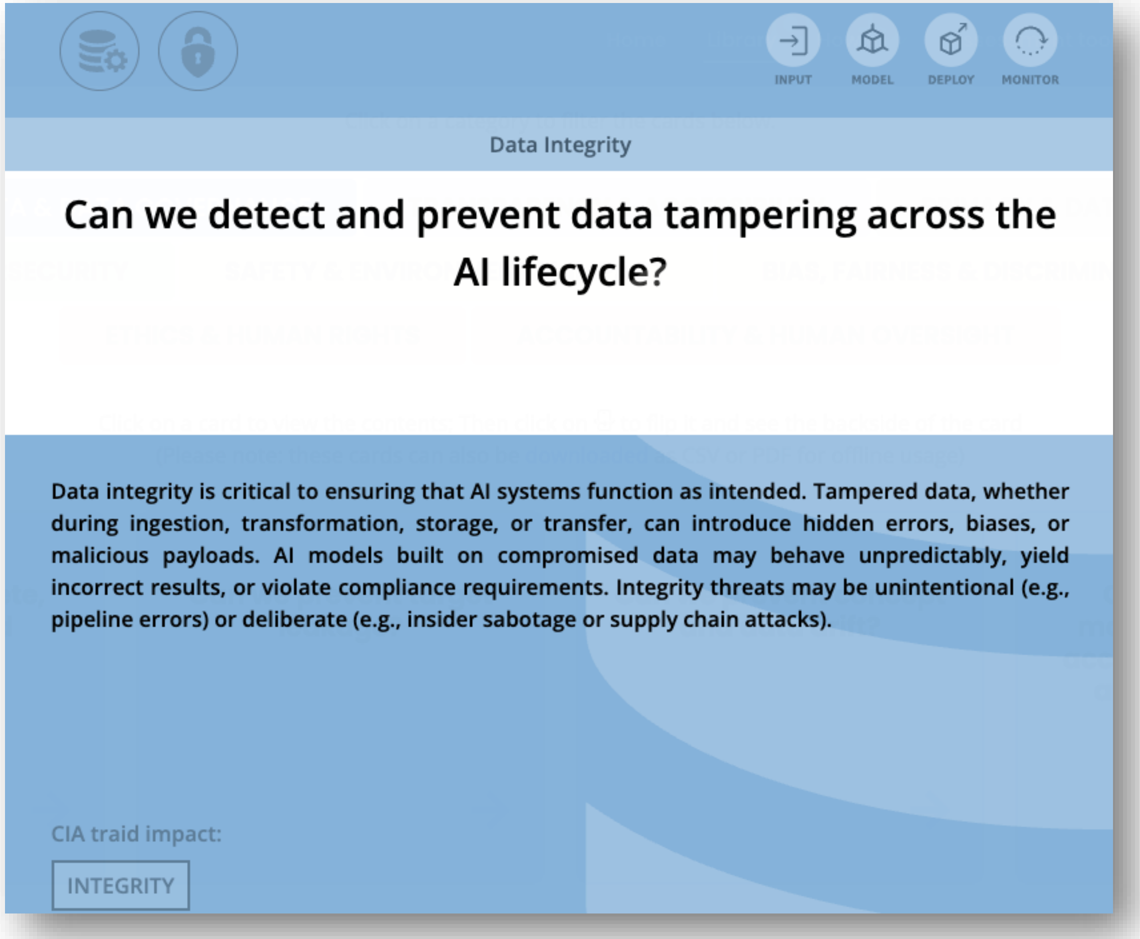
- ℹ Prevent information leakage by not revealing the existence of items in system responses.

D1 LINDDUN

Privacy threat modeling: PLOT4AI

PLOT4AI


Accountability & Human Oversight
Data & Data Governance
Ethics & Human Rights
Transparency & Accessibility
DESIGN
INPUT
MODEL
OUTPUT
DEPLOY
MONITOR
Bias, Fairness & Discrimination
Privacy & Data Protection
Safety & Environmental Impact
Cybersecurity



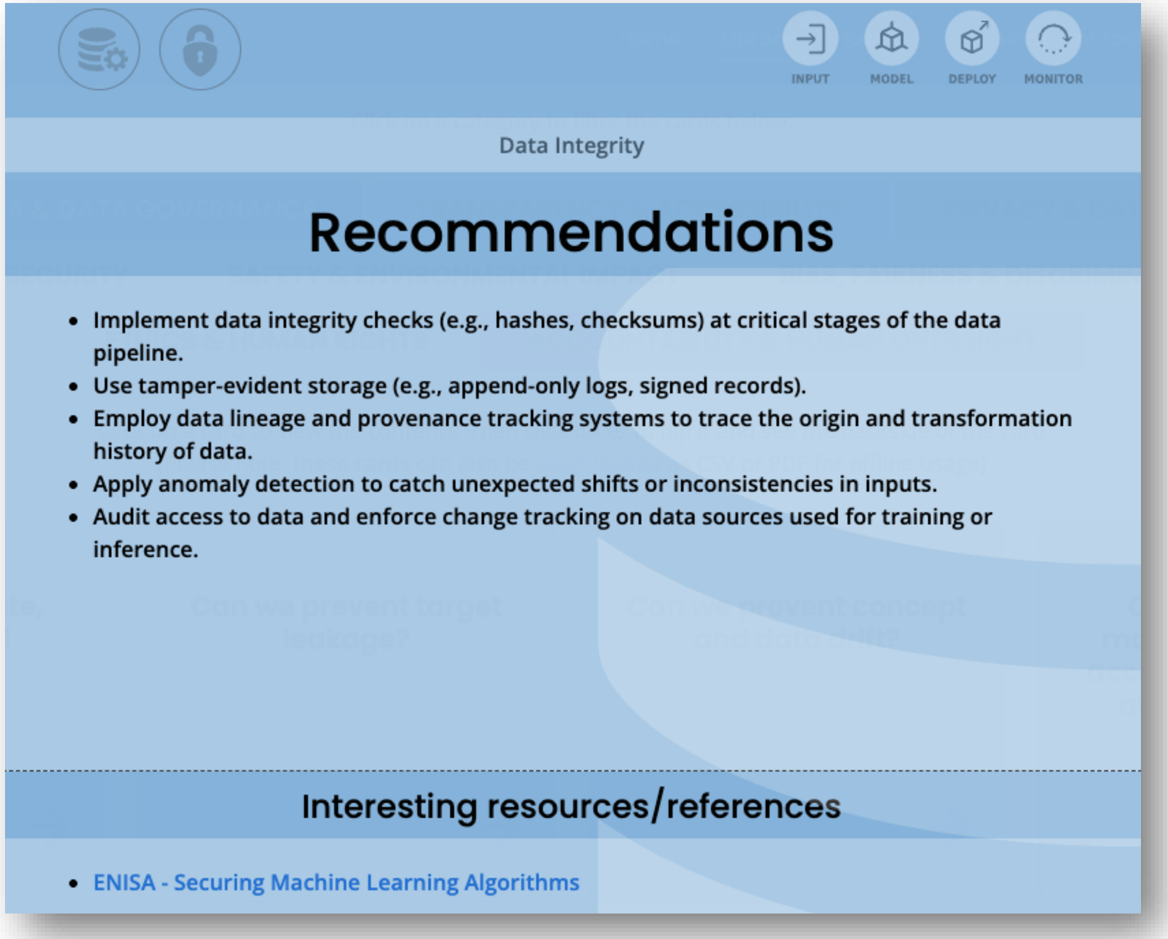
Data Integrity

Can we detect and prevent data tampering across the AI lifecycle?

Data integrity is critical to ensuring that AI systems function as intended. Tampered data, whether during ingestion, transformation, storage, or transfer, can introduce hidden errors, biases, or malicious payloads. AI models built on compromised data may behave unpredictably, yield incorrect results, or violate compliance requirements. Integrity threats may be unintentional (e.g., pipeline errors) or deliberate (e.g., insider sabotage or supply chain attacks).

CIA traid impact:

INTEGRITY



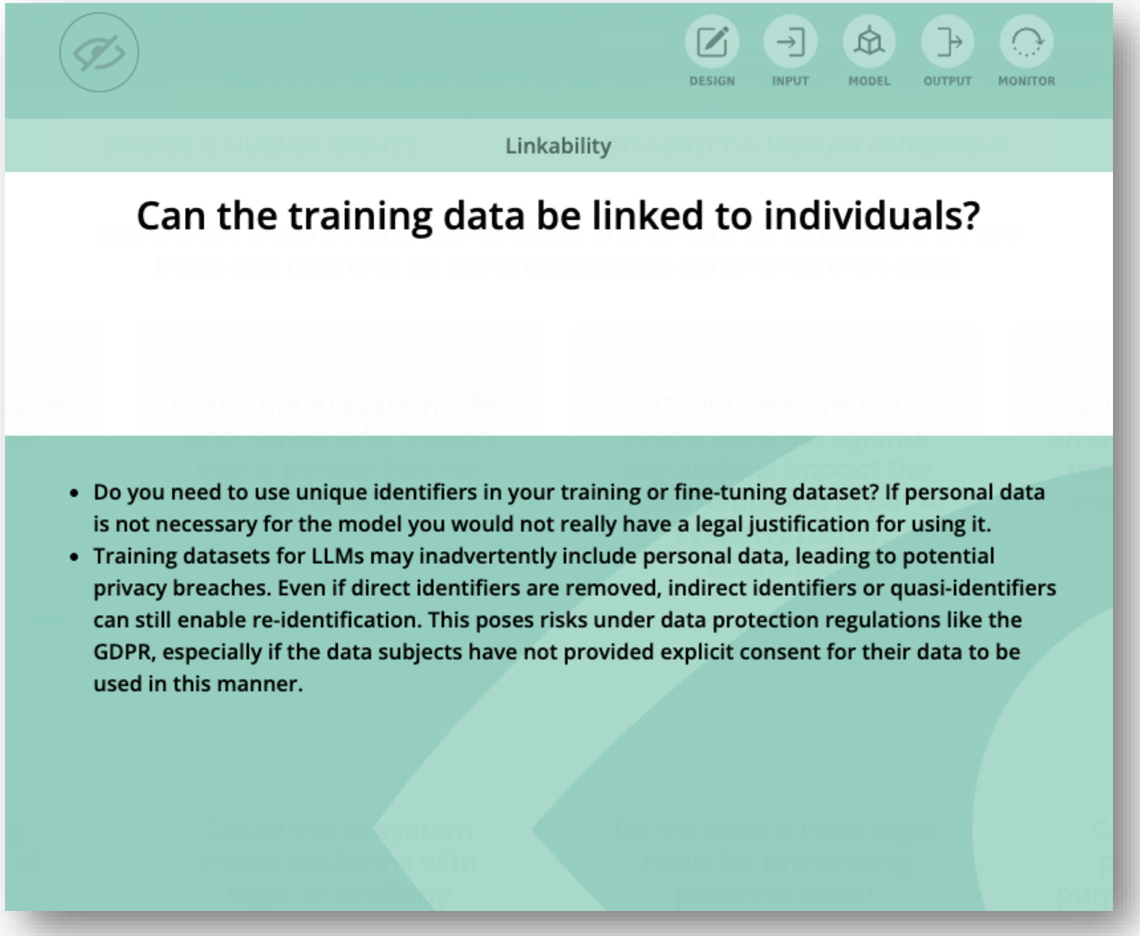
Data Integrity

Recommendations

- Implement data integrity checks (e.g., hashes, checksums) at critical stages of the data pipeline.
- Use tamper-evident storage (e.g., append-only logs, signed records).
- Employ data lineage and provenance tracking systems to trace the origin and transformation history of data.
- Apply anomaly detection to catch unexpected shifts or inconsistencies in inputs.
- Audit access to data and enforce change tracking on data sources used for training or inference.

Interesting resources/references

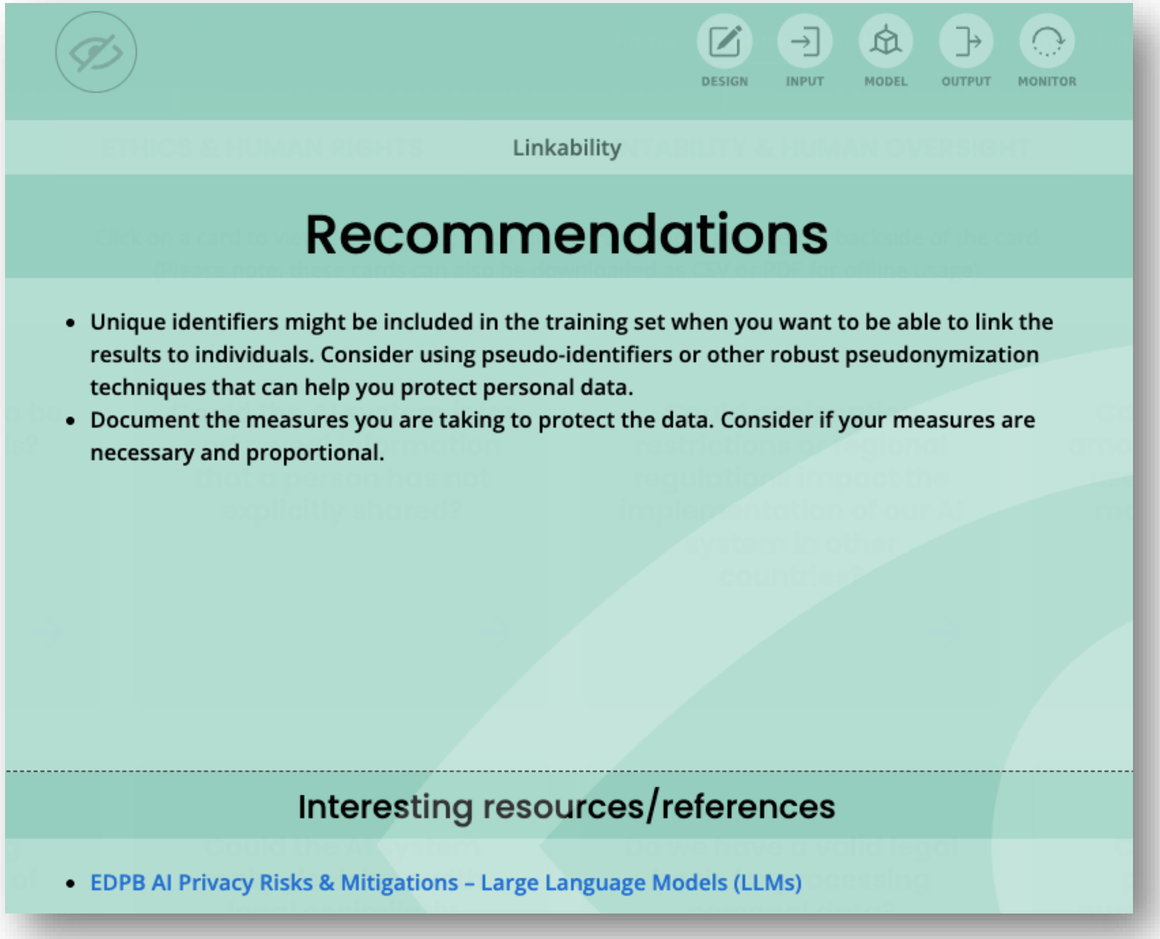
- ENISA - Securing Machine Learning Algorithms



Linkability

Can the training data be linked to individuals?

- Do you need to use unique identifiers in your training or fine-tuning dataset? If personal data is not necessary for the model you would not really have a legal justification for using it.
- Training datasets for LLMs may inadvertently include personal data, leading to potential privacy breaches. Even if direct identifiers are removed, indirect identifiers or quasi-identifiers can still enable re-identification. This poses risks under data protection regulations like the GDPR, especially if the data subjects have not provided explicit consent for their data to be used in this manner.



Linkability

Recommendations

- Unique identifiers might be included in the training set when you want to be able to link the results to individuals. Consider using pseudo-identifiers or other robust pseudonymization techniques that can help you protect personal data.
- Document the measures you are taking to protect the data. Consider if your measures are necessary and proportional.

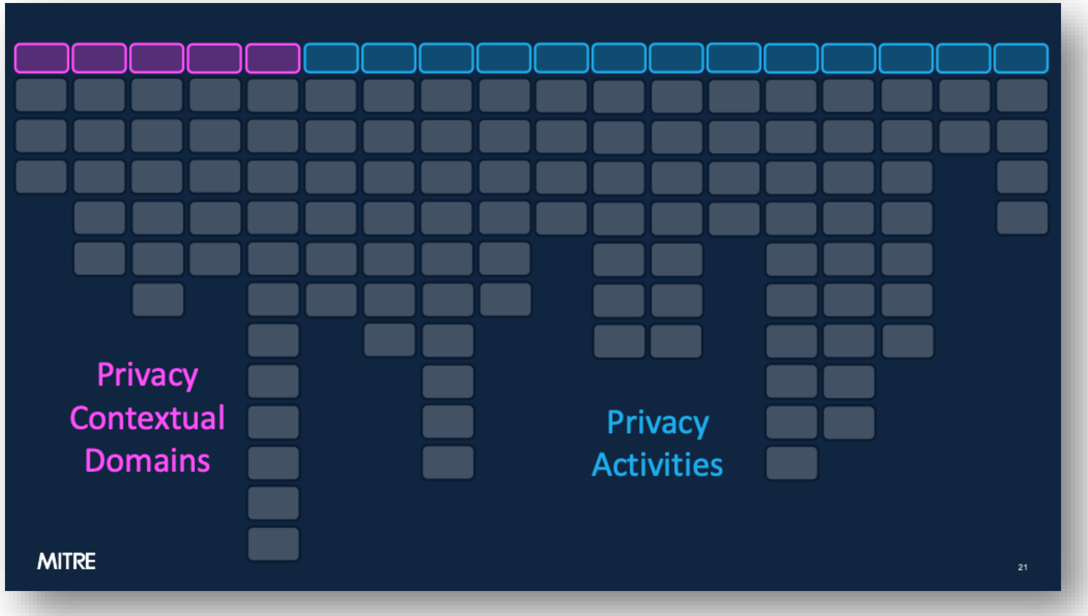
Interesting resources/references

- EDPB AI Privacy Risks & Mitigations – Large Language Models (LLMs)

Privacy threat modeling: MITRE PANOPTIC™

MITRE PANOPTIC™

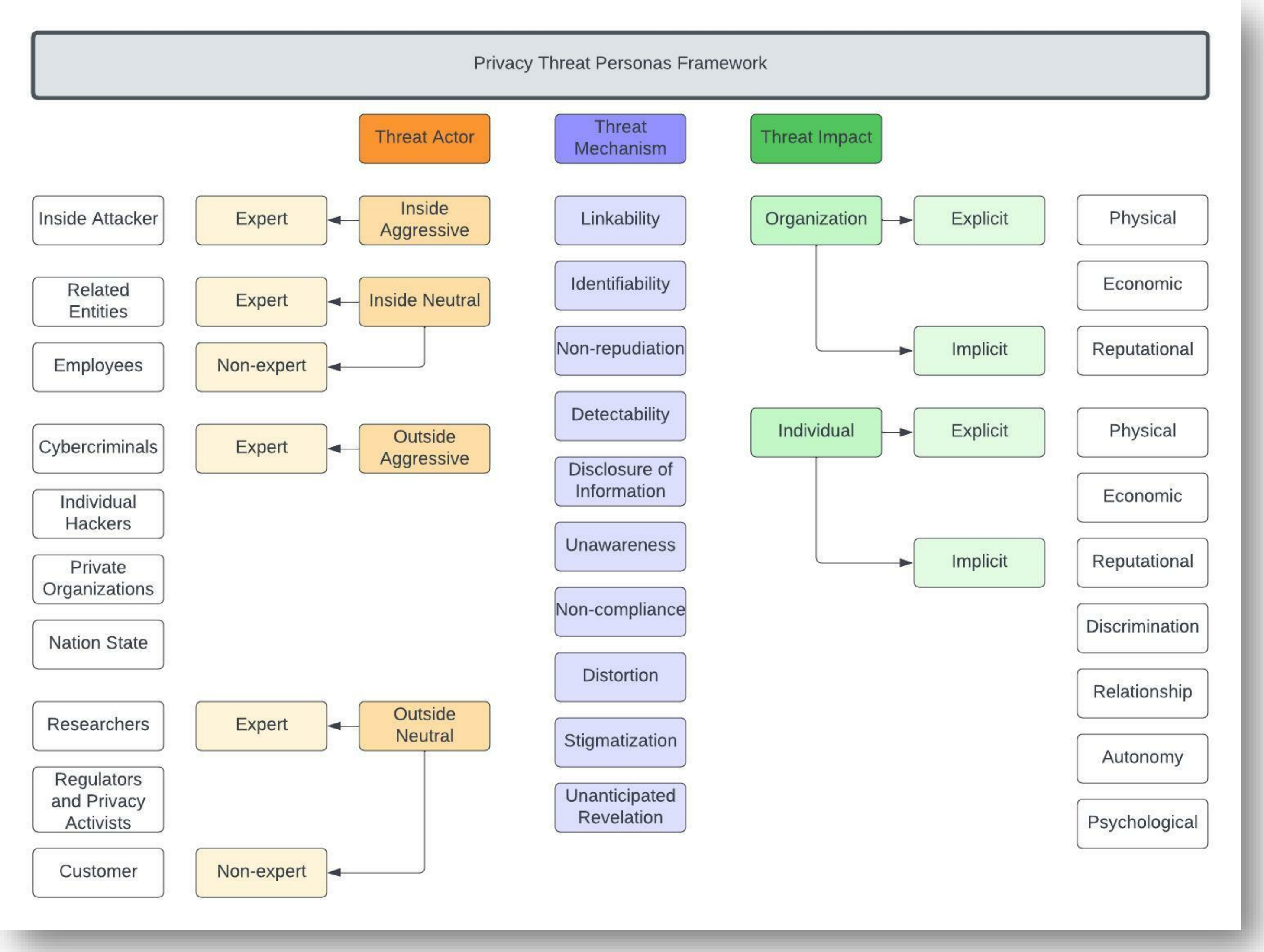
- **Privacy attack**: (in)action(s) that cause(s) a perceived privacy harm, that do(es) not solely involve cybersecurity violations
- 2x taxonomies based on 300 non-breach cases in the US
- Goal: standardized classification and 'threat language' for privacy attacks like MITRE ATT&CK® for security



Excerpt from MITRE's open training materials

Who? Why?

Comcast xCOMPASS



Privacy Threat Personas Framework

Threat Actor

Threat Mechanism

Threat Impact

Inside Attacker

Expert

Inside Aggressive

Related Entities

Expert

Inside Neutral

Employees

Non-expert

Cybercriminals

Expert

Outside Aggressive

Individual Hackers

Private Organizations

Nation State

Researchers

Expert

Outside Neutral

Regulators and Privacy Activists

Customer

Non-expert

Linkability

Identifiability

Non-repudiation

Detectability

Disclosure of Information

Unawareness

Non-compliance

Distortion

Stigmatization

Unanticipated Revelation

Organization

Explicit

Implicit

Individual

Explicit

Implicit

Physical

Economic

Reputational

Physical

Economic

Reputational

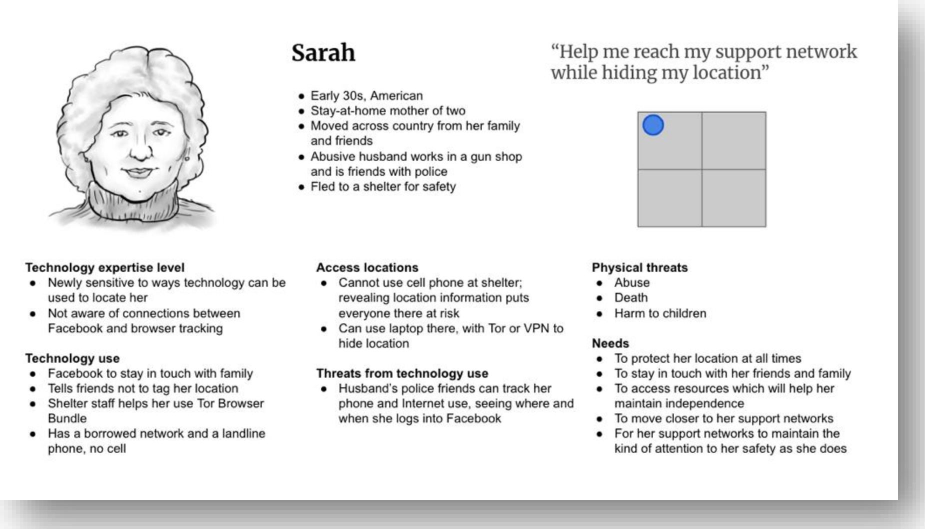
Discrimination

Relationship

Autonomy

Psychological

[Open Internet Tools Project]



Sarah

"Help me reach my support network while hiding my location"

- Early 30s, American
- Stay-at-home mother of two
- Moved across country from her family and friends
- Abusive husband works in a gun shop and is friends with police
- Fled to a shelter for safety

Technology expertise level
- Newly sensitive to ways technology can be used to locate her
- Not aware of connections between Facebook and browser tracking

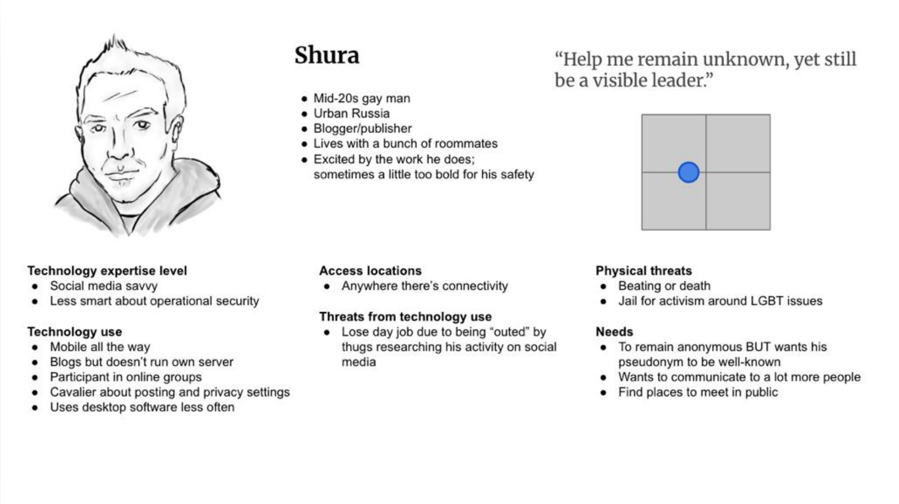
Technology use
- Facebook to stay in touch with family
- Tells friends not to tag her location
- Shelter staff helps her use Tor Browser Bundle
- Has a borrowed network and a landline phone, no cell

Access locations
- Cannot use cell phone at shelter; revealing location information puts everyone there at risk
- Can use laptop there, with Tor or VPN to hide location

Threats from technology use
- Husband's police friends can track her phone and Internet use, seeing where and when she logs into Facebook

Physical threats
- Abuse
- Death
- Harm to children

Needs
- To protect her location at all times
- To stay in touch with her friends and family
- To access resources which will help her maintain independence
- To move closer to her support networks
- For her support networks to maintain the kind of attention to her safety as she does



Shura

"Help me remain unknown, yet still be a visible leader."

- Mid-20s gay man
- Urban Russia
- Blogger/publisher
- Lives with a bunch of roommates
- Excited by the work he does; sometimes a little too bold for his safety

Technology expertise level
- Social media savvy
- Less smart about operational security

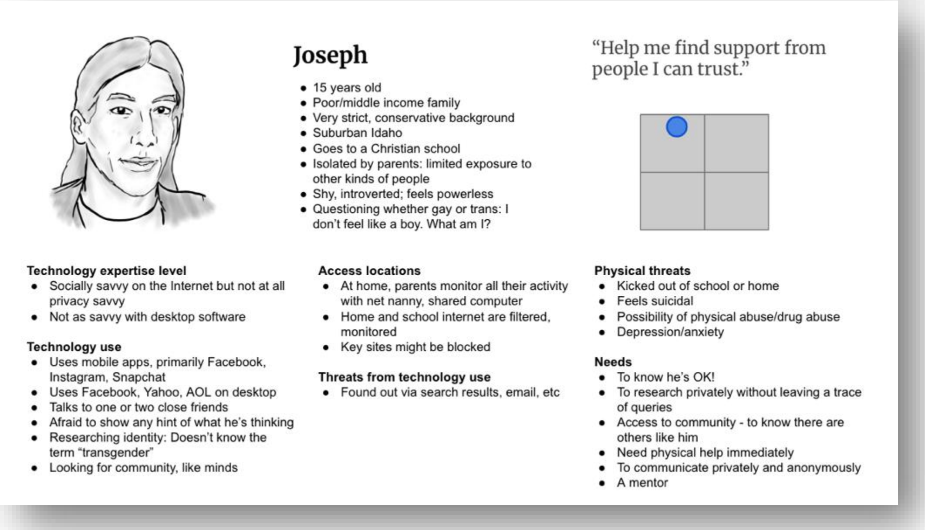
Technology use
- Mobile all the way
- Blogs but doesn't run own server
- Participant in online groups
- Cavalier about posting and privacy settings
- Uses desktop software less often

Access locations
- Anywhere there's connectivity

Threats from technology use
- Lose day job due to being "outed" by thugs researching his activity on social media

Physical threats
- Beating or death
- Jail for activism around LGBT issues

Needs
- To remain anonymous BUT wants his pseudonym to be well-known
- Wants to communicate to a lot more people
- Find places to meet in public



Joseph

"Help me find support from people I can trust."

- 15 years old
- Poor/middle income family
- Very strict, conservative background
- Suburban Idaho
- Goes to a Christian school
- Isolated by parents: limited exposure to other kinds of people
- Shy, introverted; feels powerless
- Questioning whether gay or trans: I don't feel like a boy. What am I?

Technology expertise level
- Socially savvy on the Internet but not at all privacy savvy
- Not as savvy with desktop software

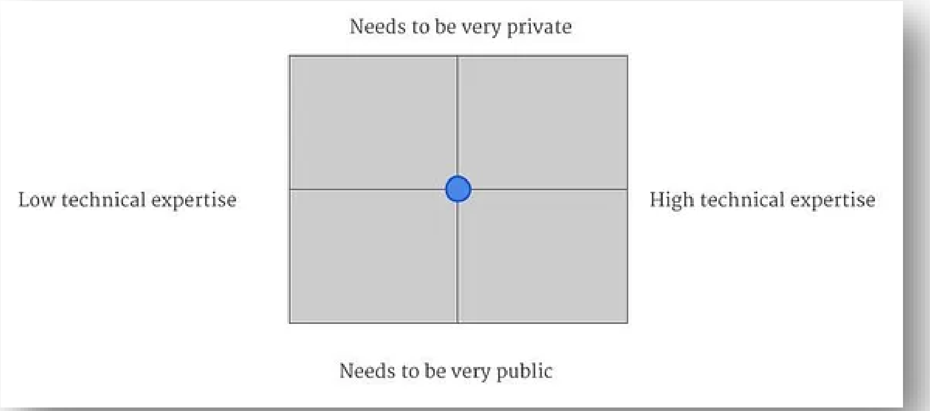
Technology use
- Uses mobile apps, primarily Facebook, Instagram, Snapchat
- Uses Facebook, Yahoo, AOL on desktop
- Talks to one or two close friends
- Afraid to show any hint of what he's thinking
- Researching identity: Doesn't know the term "transgender"
- Looking for community, like minds

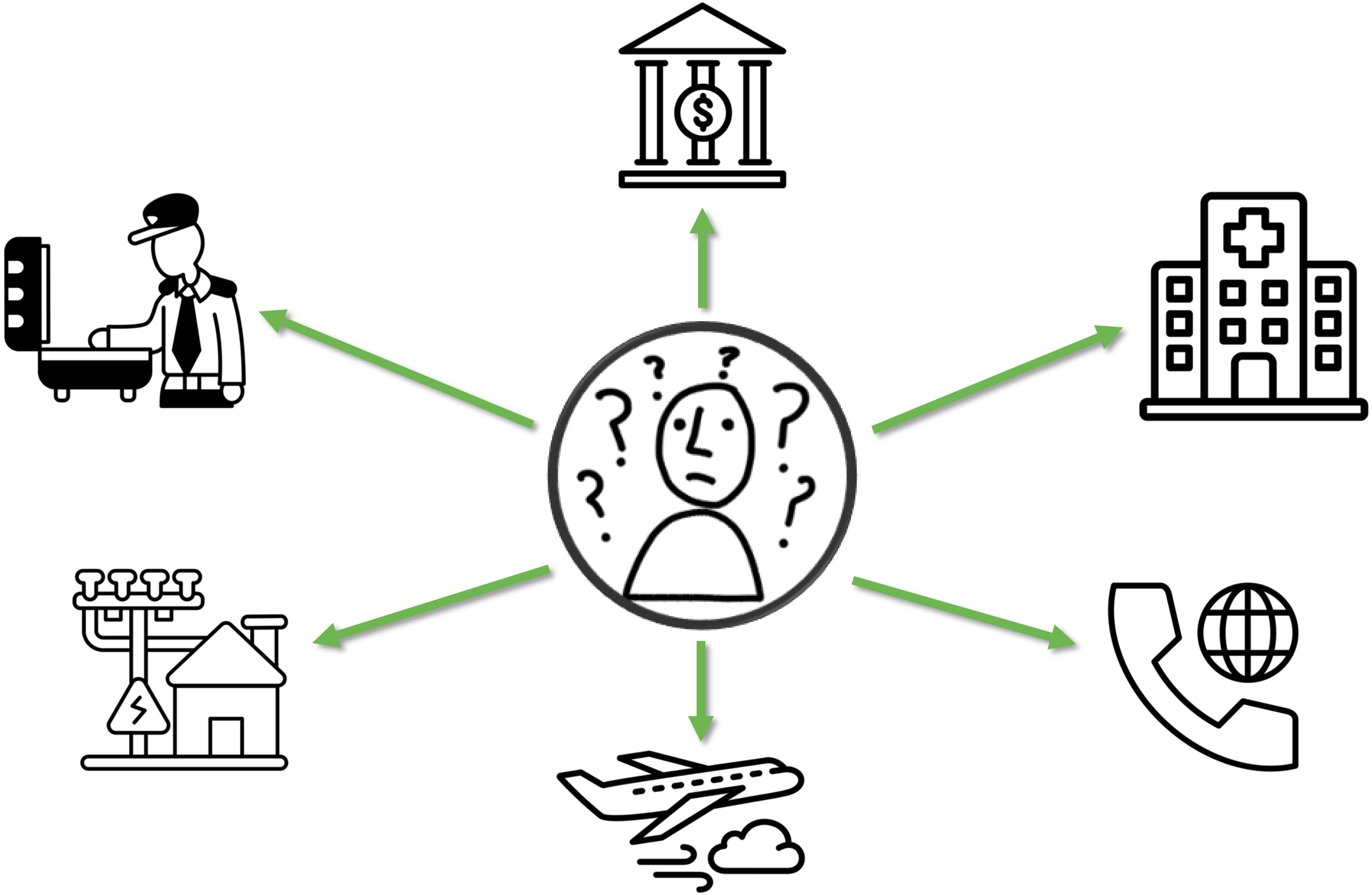
Access locations
- At home, parents monitor all their activity with net nanny, shared computer
- Home and school internet are filtered, monitored
- Key sites might be blocked

Threats from technology use
- Found out via search results, email, etc

Physical threats
- Kicked out of school or home
- Feels suicidal
- Possibility of physical abuse/drug abuse
- Depression/anxiety

Needs
- To know he's OK!
- To research privately without leaving a trace of queries
- Access to community - to know there are others like him
- Need physical help immediately
- To communicate privately and anonymously
- A mentor

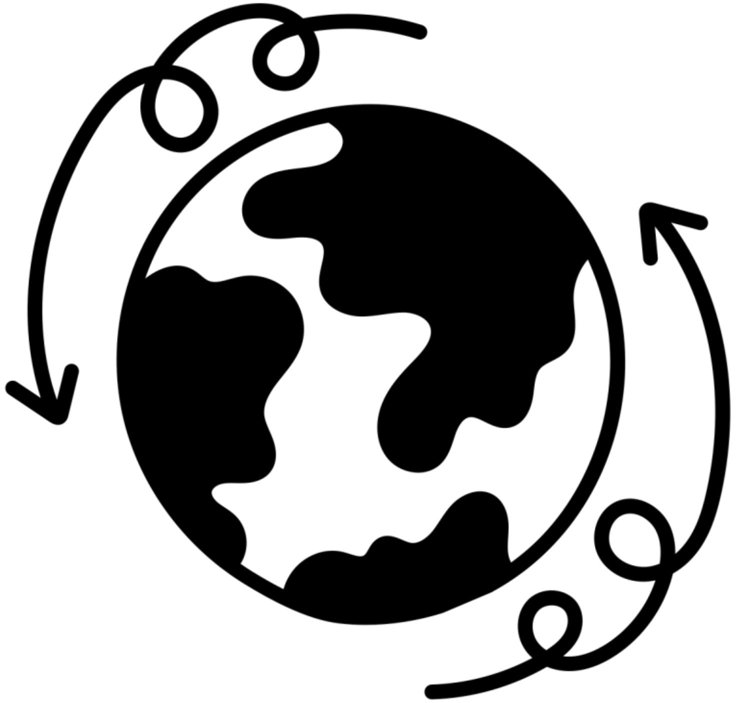




Example: IP address transfer to Boggleland

Recap

- Your company is acquiring a company that provides encrypted email services
- Integrating their product with your SSO will mean that timestamped EU customer IP addresses are transferred to **Boggleland** on each login
- You are conducting a DPIA and are unsure how to determine:
  - What could go wrong for data subjects?
  - How bad would it be (for us and them!)?
  - Should we do anything about it, and if so what?



Data transferred

```
2025-10-28T06:32:45.123Z [INFO] User login:
uuid=123e4567-e89b-12d3-a456-426614174000,
ip_address= 83.164.100.0, event=login_success
```

Who? Why?

What might go wrong – any ideas?

Who? Why?

xCompass potential attacker: nation state with legal power to access our SSO logs in Boggleland and ISP logs

Vulnerable (legitimate) users:

- Journalists and their sources; lawyers and their clients
- Government officials; activists
- Other high-risk categories depending on political context (potentially reprisals based on religion, political affiliation, sexual orientation, physical or mental health, ...)

What? How? - MITRE PANOPTIC™

- **Environment:** digital (POC01.01)
- **Distribution:** one to one (PC02.02)
- **Interaction:** ongoing interaction (PC03.01.03)
- **Engagement**: populations with sensitive characteristics (PC04.01.02-08 + 10-11): *race & ethnicity, political opinion, religious and philosophical beliefs, gender, sexual orientation & gender identity, sex life, genetics, illness or injury, other context-specific populations*
- **Data type:** persistent direct identifier (PC05.15.01), persistent pseudo-identifier (PC05.15.02), location (PC05.01)

What? How? - MITRE PANOPTIC™

Potential privacy threat actions by us:

- **Notice and consent:** absent, no opt-out
- **Collection:** application or device use, tracking & affording tracking
- **Identification:** identifier assignment, implicit identification
- **Manageability:** no individual control of information disclosure
- **Sharing:** affording revelations
- **Deviations:** deviating from stated policy, deviating from regulatory requirements

What? How? - MITRE PANOPTIC™

Potential privacy threat actions by the nation state:

- **Processing:** deriving new information, behavioral analysis
- **Use:** implication, targeting, intrusion, reprisal

What? How? – PLOT4AI

From the Privacy & Data Protection category:

- Are we using metadata that could reveal personal data or behavior patterns? *Yes*
- Are we processing special categories of personal data or sensitive data? *No, but the fact someone uses the service is revealing*
- Are we transferring personal data to countries that lack adequate privacy protections? *Yes*
- Are we able to comply with all the applicable GDPR data subject rights? *Maybe*

What? How? - LINDDUN

- **Linking** IP address locations for a specific user over time to potentially **identify** them or **detect** that a target does indeed use the service
- Leading to **non-repudiation** of the fact the user used the service
- **Unawareness & unintervenability:** without sufficient notice, vulnerable users will be unable to protect themselves
- **Non-compliance:** cross-border data transfer (basis?), notice, consent

Case study

Time to split into groups! Check your handouts for the case study

Impact and likelihood

Risk evaluation is hard!

Likelihood	Definition 1 (data protection !!)	Definition 2 (security)	Definition 3 (government)
Rare	Has never occurred	Once in 10 years	Once in 100 years
Unlikely	Annually	Once in 5 years	Once in 25 years
Possible	Monthly	Once in 2 years	Once in 10 years
Likely	Weekly	Once a year	Once in 3 years
Certain	Almost daily	Multiple times a year	Once a year

Agree on definitions for likelihood and impact that makes sense in your organization's context!

Risk evaluation in context: Catalan DPA’s FRIA model (2025)

Risk matrix: effort to overcome the prejudice and to reverse adverse effects

Level	Definition
Low	Suffered prejudice can be overcome without any problem (e.g. time spent amending information, annoyances, irritations, etc.)
Medium	A few difficulties (e.g. extra costs, fear, lack of understanding, stress, minor physical ailments, etc.)
High	Serious difficulties (e.g. economic loss, property damage, worsening of health, etc.)
Very high	May not be overcome (e.g. long-term psychological or physical ailments, death, etc.)

Risk evaluation in context: Catalan DPA’s FRIA model (2025)

FRIA model

- Impact = (gravity * effort to overcome the prejudice and reverse its effects)
- Likelihood = (probability * exposure)
- Overall risk = (gravity * effort to overcome the prejudice and reverse its effects) * (probability * exposure)

Risk evaluation in context: Catalan DPA’s FRIA model (2025)

FRIA model

- Prejudice X is unlikely to occur *(low probability)*
- But the majority of the identified population would be affected *(high exposure)*
- Minor prejudice encountered by affected individuals/groups *(low gravity)*
- Suffered prejudice could be overcome with a few difficulties such as stress and extra cost *(medium effort)*
- Overall risk = high likelihood * medium severity = **high risk**

Q&A