



# DSC 291 Privacy-sensitive Data Systems (week 5a)

Haojian Jin

# Logistics

1. Final project
  1. Abstract due Feb. 15.
  2. Final report due Mar. 19
2. Grades
  1. Would be more strict.

# Recap: Policies

- Systems
  - P3P
  - Peekaboo
  - Hippocratic database
- Privacy principles
  - Interfaces for policies, auditors, developer incentives, machine-readable, usability, default options

# Privacy Reviews and Compliance

- Why
- Privacy by Design
- Practices in the industry
- Privacy Reviews



3,652,539 views | Feb 16, 2012, 11:02am

## How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did



Kashmir Hill Former Staff

Tech

Welcome to The Not-So Private Parts where technology & privacy collide

⌚ This article is more than 2 years old.

Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy. Target [TGT +0%](#), for example, has figured out how to data-mine its way into your womb, to figure out whether you have a baby on the way long before you need to start buying diapers.



Target has got you in its aim

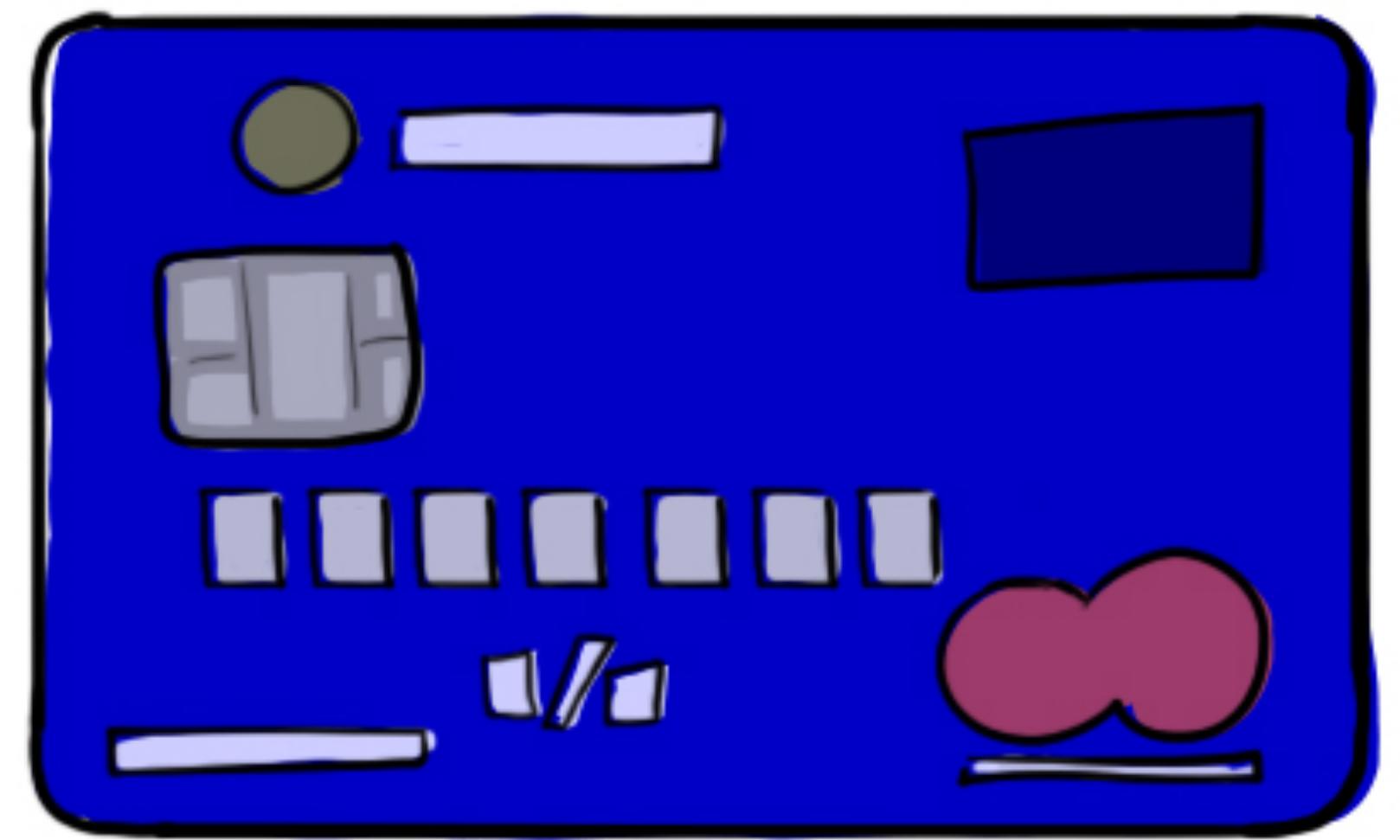
**Target predicted that a teenage girl might be pregnant and sent a diaper coupon to the girl.**

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did.

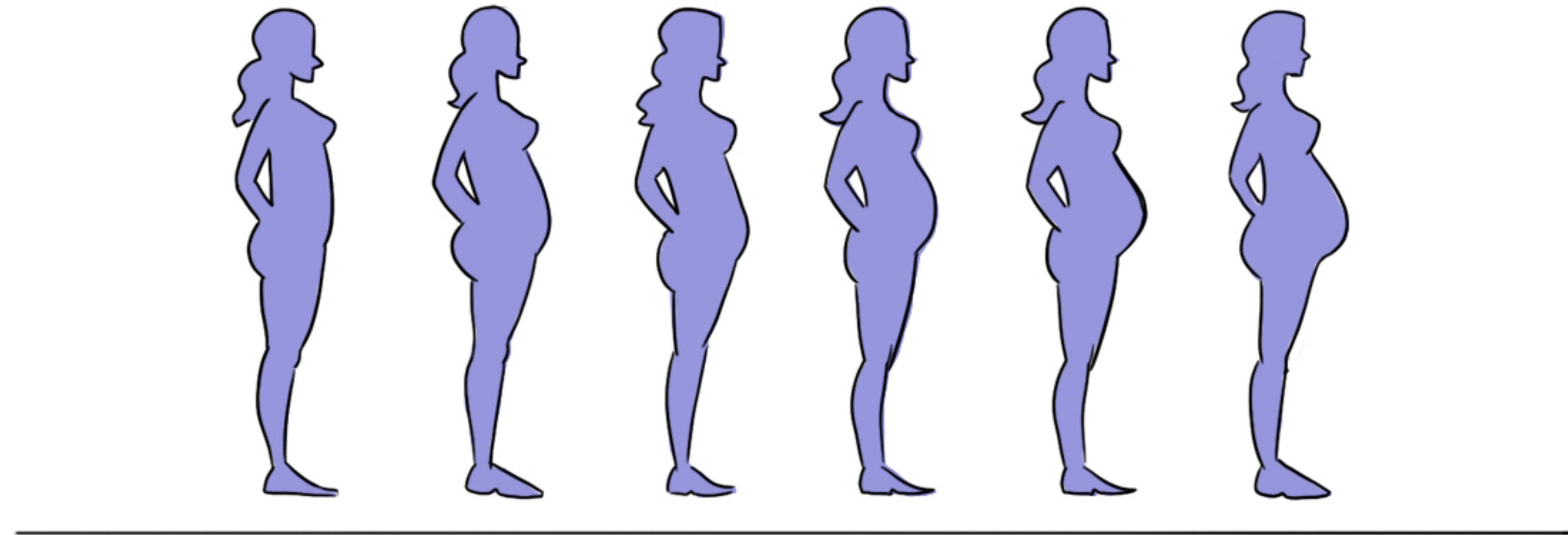
Kashmir Hill, Forbes, 2012



Target assigns every customer a guest ID number, tied to their credit card, name, or email address and collects their demographic information through various channels.



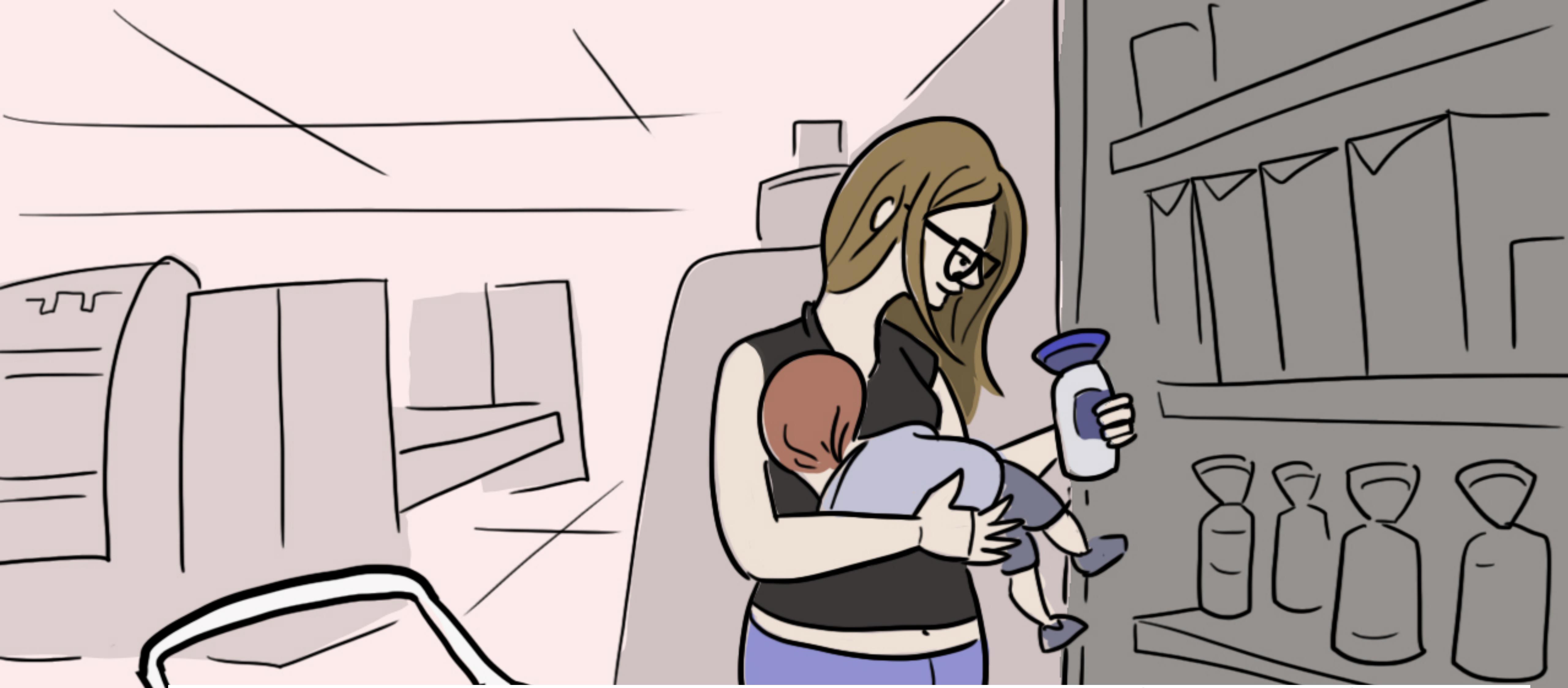
Using the guest ID, Target tracks everything the customer purchased in the past and develops sophisticated models to improve their business.



Pregnant women purchase different items during the gestation period. For example, sometime in the first 20 weeks, pregnant women loaded up on supplements like calcium, magnesium and zinc.



Target will send coupons for baby items to customers according to their pregnancy predictions.



The reason behind it is that research shows that frustrated first time parents are more likely to change their lifelong shopping behavior during the overwhelming period.

Shop Target for everything  
you need to need!



Target tries to hook parents-to-be at that crucial moment before they turn into rampant — and loyal — buyers of all things pastel, plastic, and miniature.

# Postmortem

JOHN LUNNEY, SUE LUEDER, AND GARY O'CONNOR, GOOGLE | APRIL 24, 2018

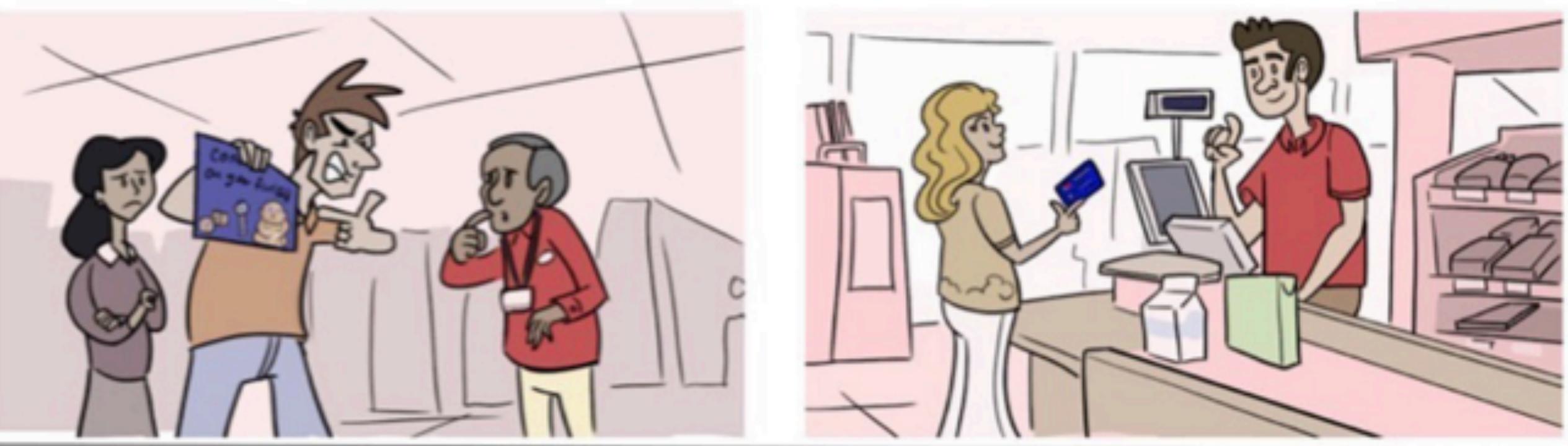


## Postmortem culture: how you can learn from failure

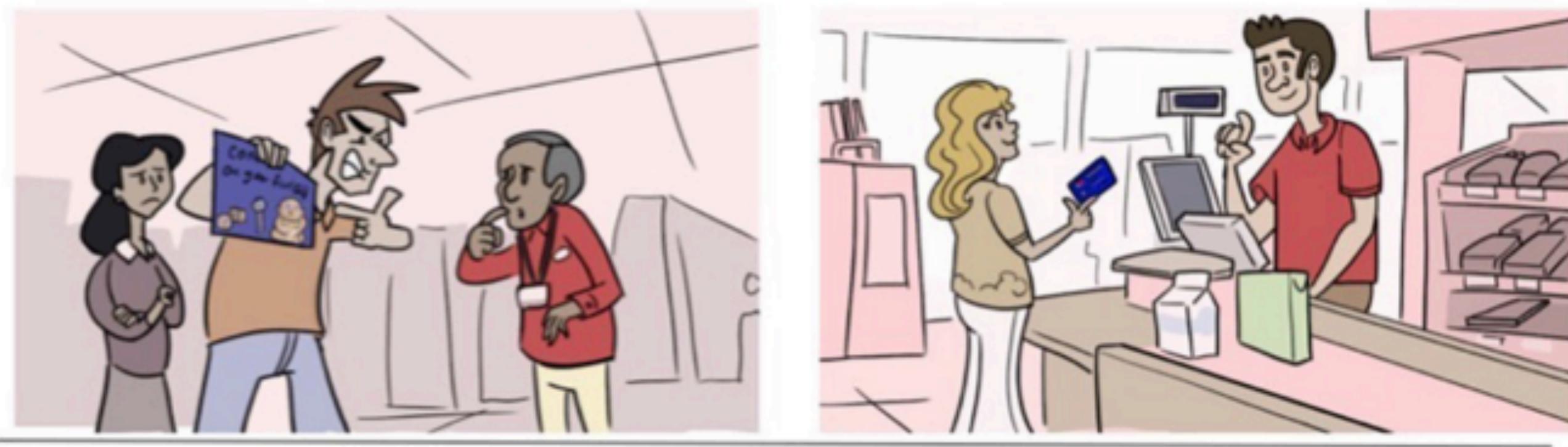


Failures are an inevitable part of innovation and can provide great data to make products, services, and organizations better. Google uses “postmortems” to capture and share the lessons of failure.

# Which step is wrong?



# Privacy by Architecture, Policies, Design, Default ...



# Privacy by policy vs. architecture

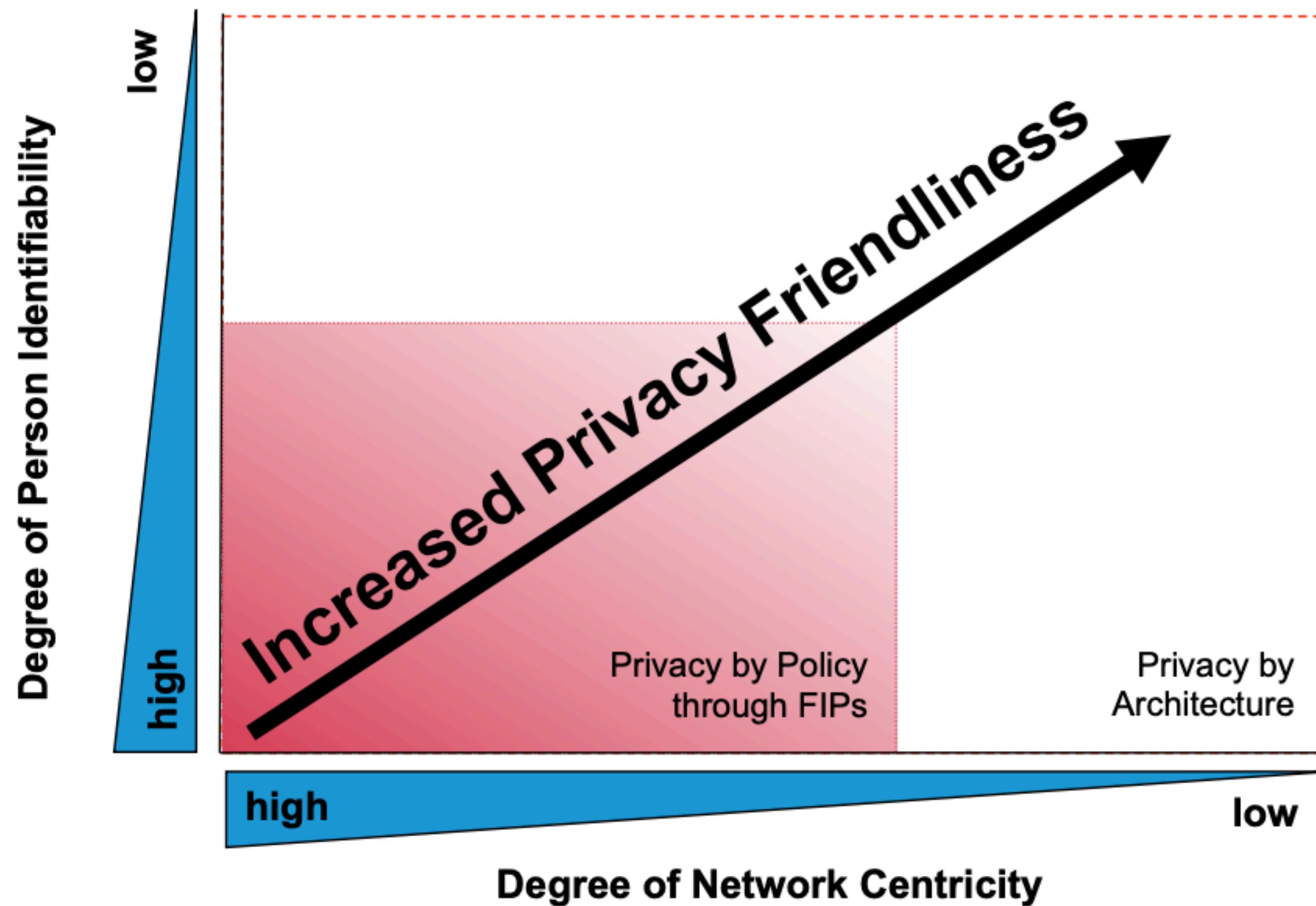
## Privacy by Policy

- Through laws and policies
- Requires enforcement, tech can facilitate compliance
- Violations possible due to bad actors, mistakes, government mandates

## Privacy by Architecture

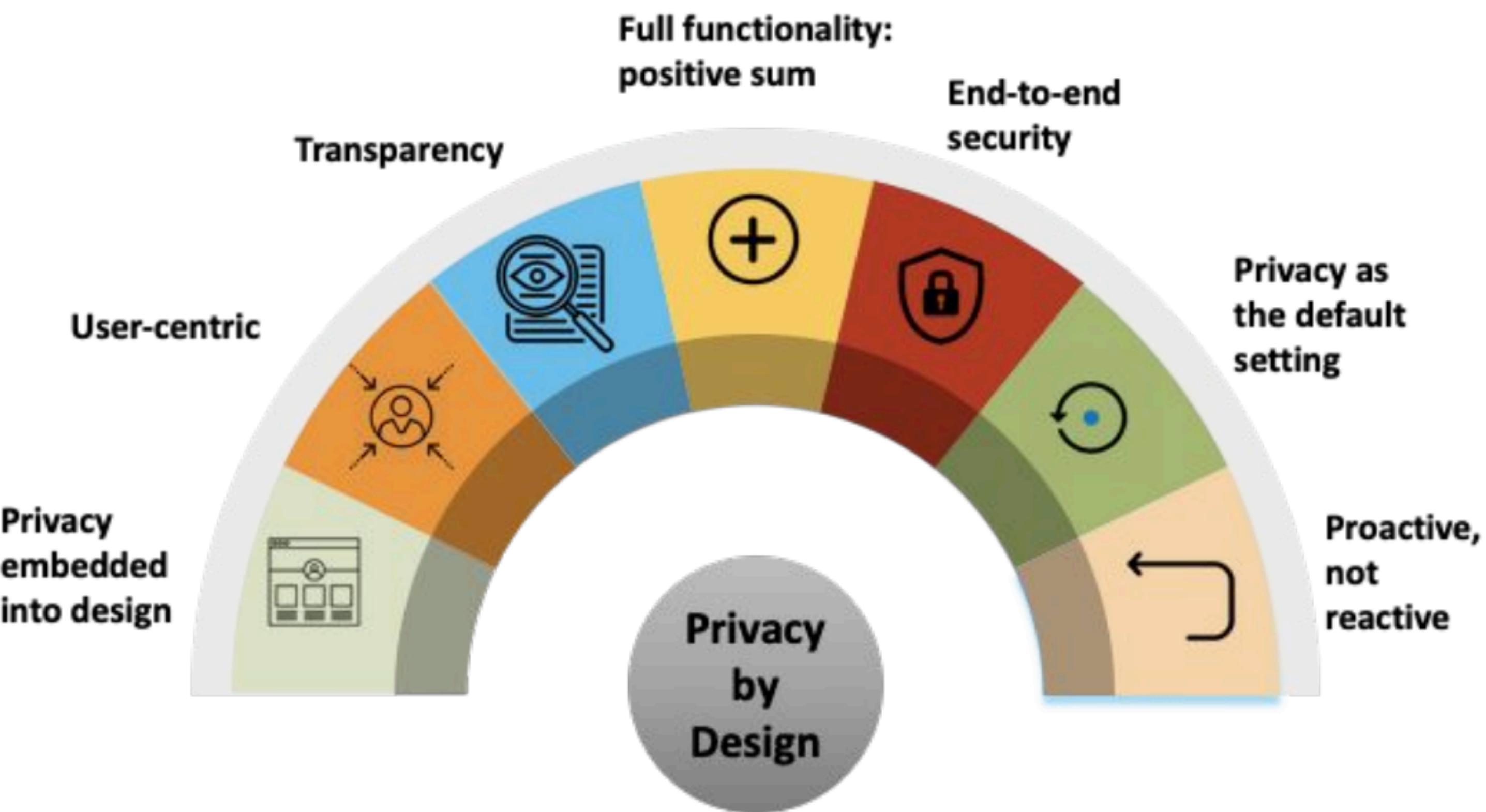
- Through technology
- Reduces need to rely on trust & external enforcement
- Violations possible tech fails
- May be viewed as too expensive or restrictive

# What system features tend to lead to more or less privacy?



<b>Privacy stages</b>	<b>Identifiability</b>	<b>Approach to privacy protection</b>	<b>Linkability of data to personal identifiers</b>	<b>System Characteristics</b>
0	identified	privacy by policy (notice and choice)	linked	<ul style="list-style-type: none"> <li>• unique identifiers across databases</li> <li>• contact information stored with profile information</li> </ul>
1	pseudonymous	privacy by architecture	linkable with reasonable & automatable effort	<ul style="list-style-type: none"> <li>• no unique identifiers across databases</li> <li>• common attributes across databases</li> <li>• contact information stored separately from profile or transaction information</li> </ul>
2		privacy by architecture	not linkable with reasonable effort	<ul style="list-style-type: none"> <li>• no unique identifiers across databases</li> <li>• no common attributes across databases</li> <li>• random identifiers</li> <li>• contact information stored separately from profile or transaction information</li> <li>• collection of long term person characteristics on a low level of granularity</li> <li>• technically enforced deletion of profile details at regular intervals</li> </ul>
3	anonymous	privacy by architecture	unlinkable	<ul style="list-style-type: none"> <li>• no collection of contact information</li> <li>• no collection of long term person characteristics</li> <li>• <math>k</math>-anonymity with large value of <math>k</math></li> </ul>

# Privacy By Design (since 1995)



**Redesigning IP Geolocation:  
Privacy by Design and Online  
Targeted Advertising**



October 2010

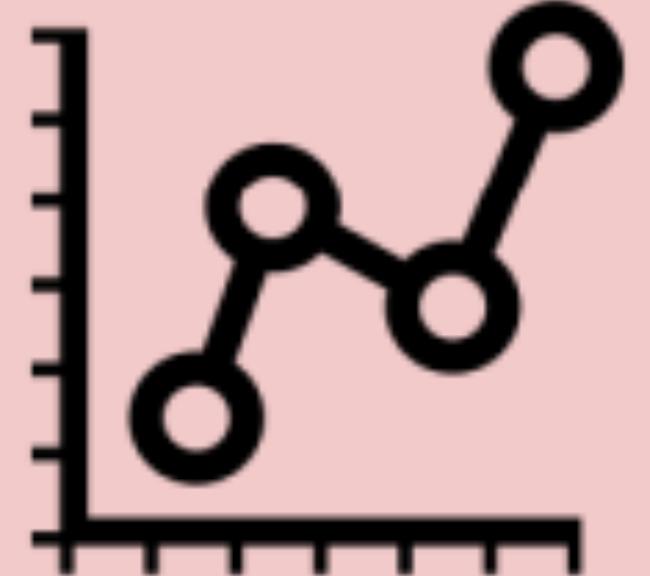
**P**  
Ann Cavoukian, Ph.D.  
Information and Privacy Commissioner,  
Ontario, Canada

With contributions from:  
**beringmedia** ::::

# Challenges of Privacy By Design

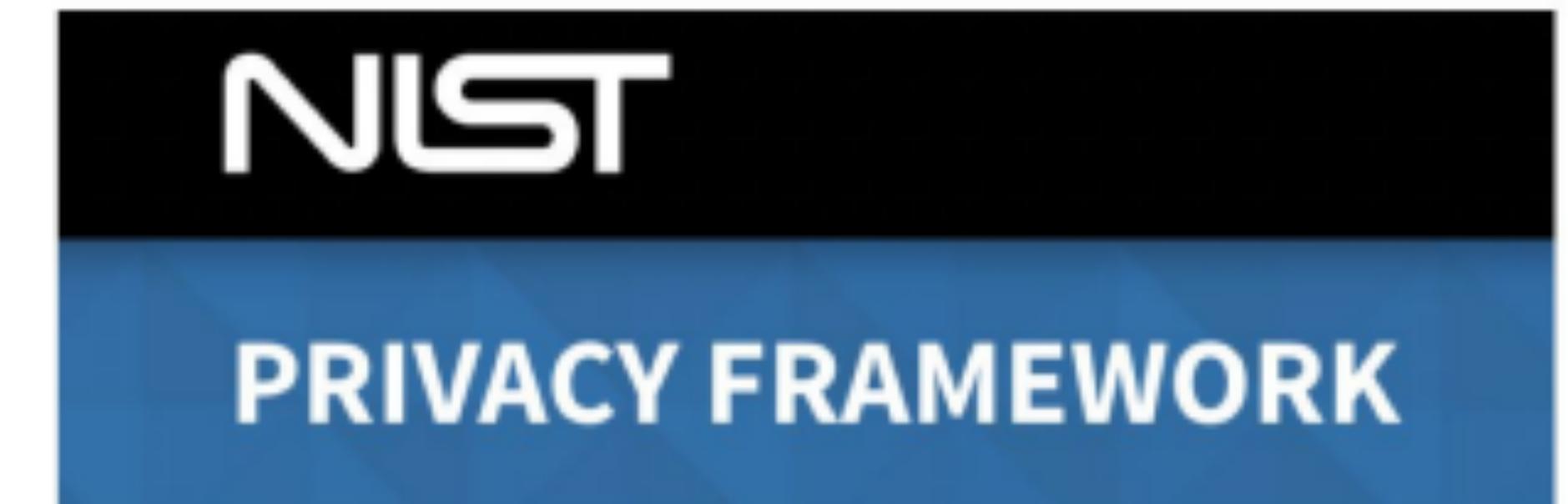
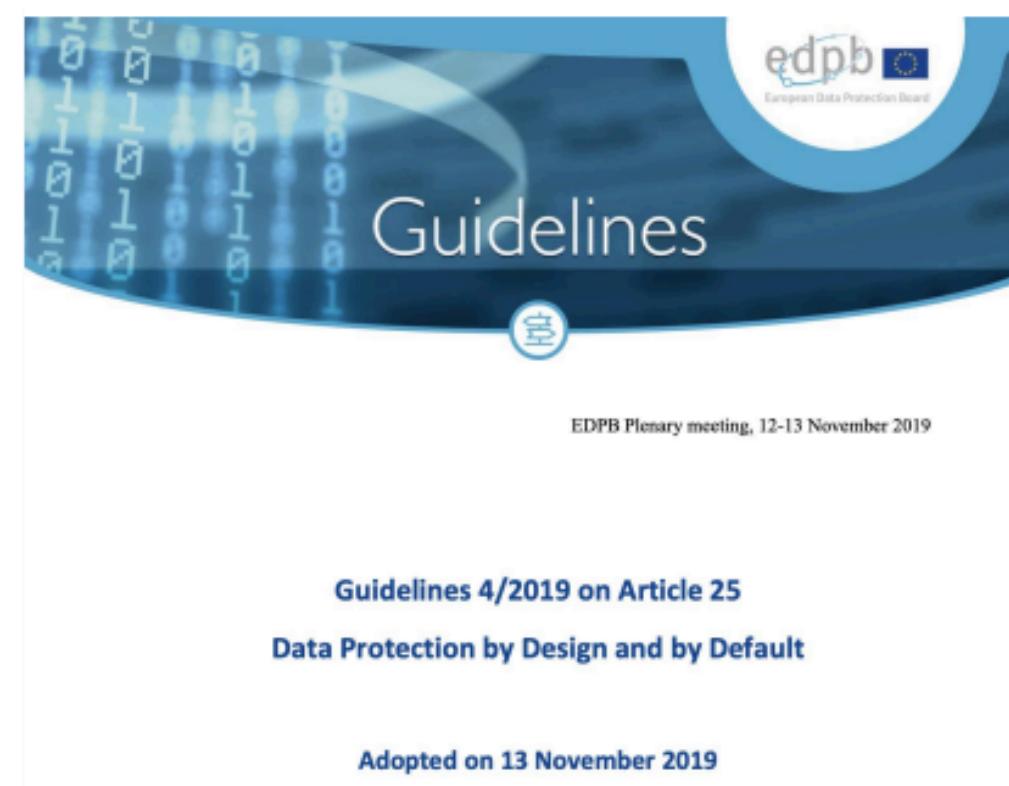
- **Vague**
- **Not practical**
- **Lack incentives**
- **Expensive**

# GDPR Art 25: Which factors?

<p>Appropriate Measures and Principles</p> 	<p>State of the art</p> 	<p>Nature, scope, context and purpose</p> 
<p>Effective (Demonstrate with metrics)</p> 	<p>Costs</p> 	<p>Risk for the individual</p> 

# When to implement the controls?

- Decision making on the tech stack, vendor
- Abstract, engineering design document, prototypes
- Implementation is effective
- Periodic vendor reviews
- Data breaches
- Data deletion



# Privacy Impact Assessment



U.S. DEPARTMENT OF AGRICULTURE

GLOSSARY ASKUSDA RECALLS CONTACT US

HOME

TOPICS

OUR AGENCY

PRIORITIES

MEDIA



Help us improve  
USDA.gov

## Privacy Impact Assessments



Privacy Impact Assessments

OSSPI

# PRIVACY IMPACT ASSESSMENTS



06:51



A **privacy impact assessment** (PIA) is an analysis of how personally identifiable information (PII) is handled to ensure **compliance** with appropriate regulations, determine the privacy risks associated with **information systems or activities**, and evaluate ways to reduce the privacy risks. A PIA is both an analysis and a formal document detailing the process and the outcome of the analysis.

Below is a list of the USDA's publicly available PIAs:

### Mission Areas:

# Privacy Impact Assessment

A methodology for

- assessing the impacts on privacy of a project, policy, program, service, product, or other initiative which involves the processing of personal information and,
- in consultation with stakeholders, for taking remedial actions as necessary in order to avoid or minimize negative impacts

# PIA is a process

- Should begin at early stages of a project
- Should continue to end of project and beyond

# Privacy Analysis Worksheet

## SECTION A

### **Summary Information**

1. Name of project or system:  
<Please enter the project or system name here.>
2. Description of project or system and its purpose:  
<Please provide a general description of the project or system, and its purpose using a non-technical description, if statutory, provide citation.>
3. Contact Name, Title, Telephone Number and Organization:  
<Please provide information here.>

### **Specific Questions**

1. Does this project or system collect, maintain, retrieve or share personal information that can be used to directly or indirectly identify an individual?

- NO. A PIA is not required for this project. Skip to Signature Page.  
 YES. A PIA is required for this project.

<Please provide a specific description of the information that might be collected or maintained.>

2. Does this project or system retrieve information using a personal identifier?

- NO. A Privacy Act SORN is not required for this project. Skip to Signature Page.  
 YES. A Privacy Act SORN is required for this project.

<Please provide a description of the data fields that might be used to retrieve the information.>

Is there an existing Privacy Act System of Records Notice (SORN)?

- NO. <Contact [privacyhelp@sec.gov](mailto:privacyhelp@sec.gov) for assistance.>  
 YES. The existing SORN may need to be modified to reflect changes.  
<Please provide the system notice number.>

# Why carry out a PIA?

## To manage risks

- Negative media attention
- Reputation damage
- Legal violations
- Fines, penalties
- Privacy harms
- Opportunity costs

## To derive benefits

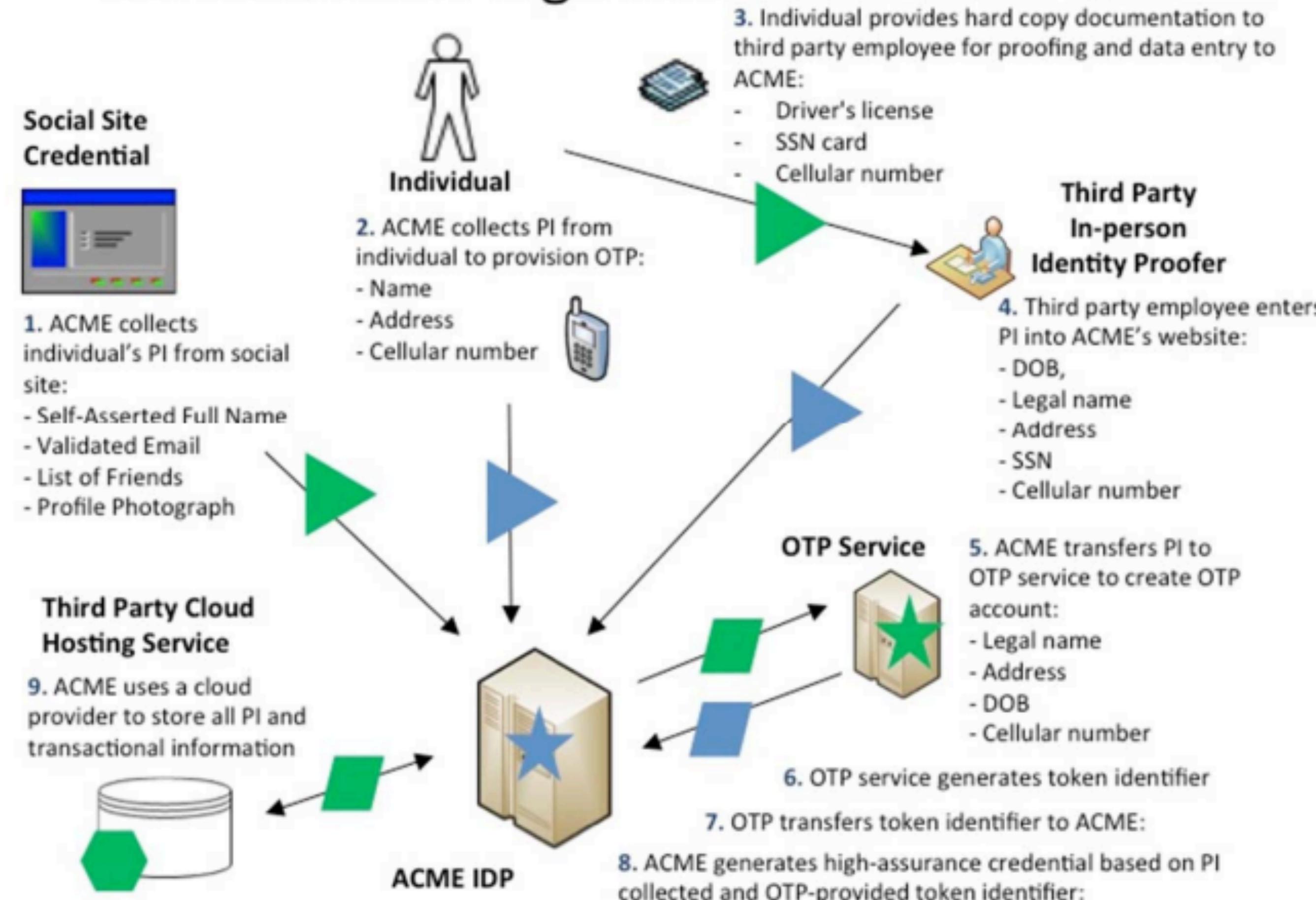
- Increase trust
- Avoid future liability
- Early warning system
- Facilitate privacy by design early in design process
- Enforce or encourage accountability

# Who has to carry out PIAs?

- US administrative agencies, when developing or procuring IT systems that include PII
  - Required by E-Government Act of 2002
  - Government agencies in many other countries
  - Sometimes done by private sector
    - Case studies from Vodafone, Nokia, and Siemens in PIA book

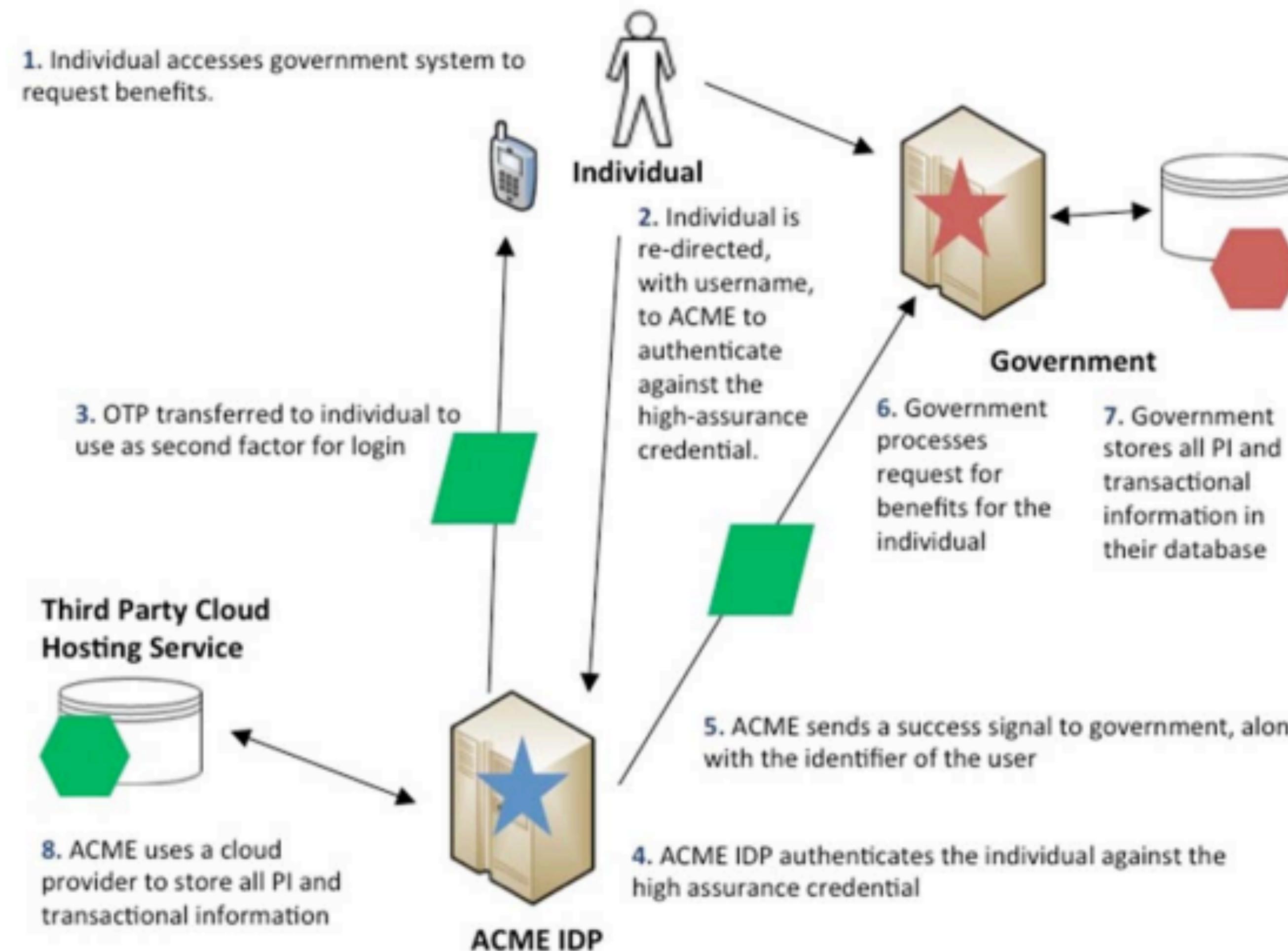
# PRAM - Task 1 - Map data processing within the system

## Generation of high-assurance credential

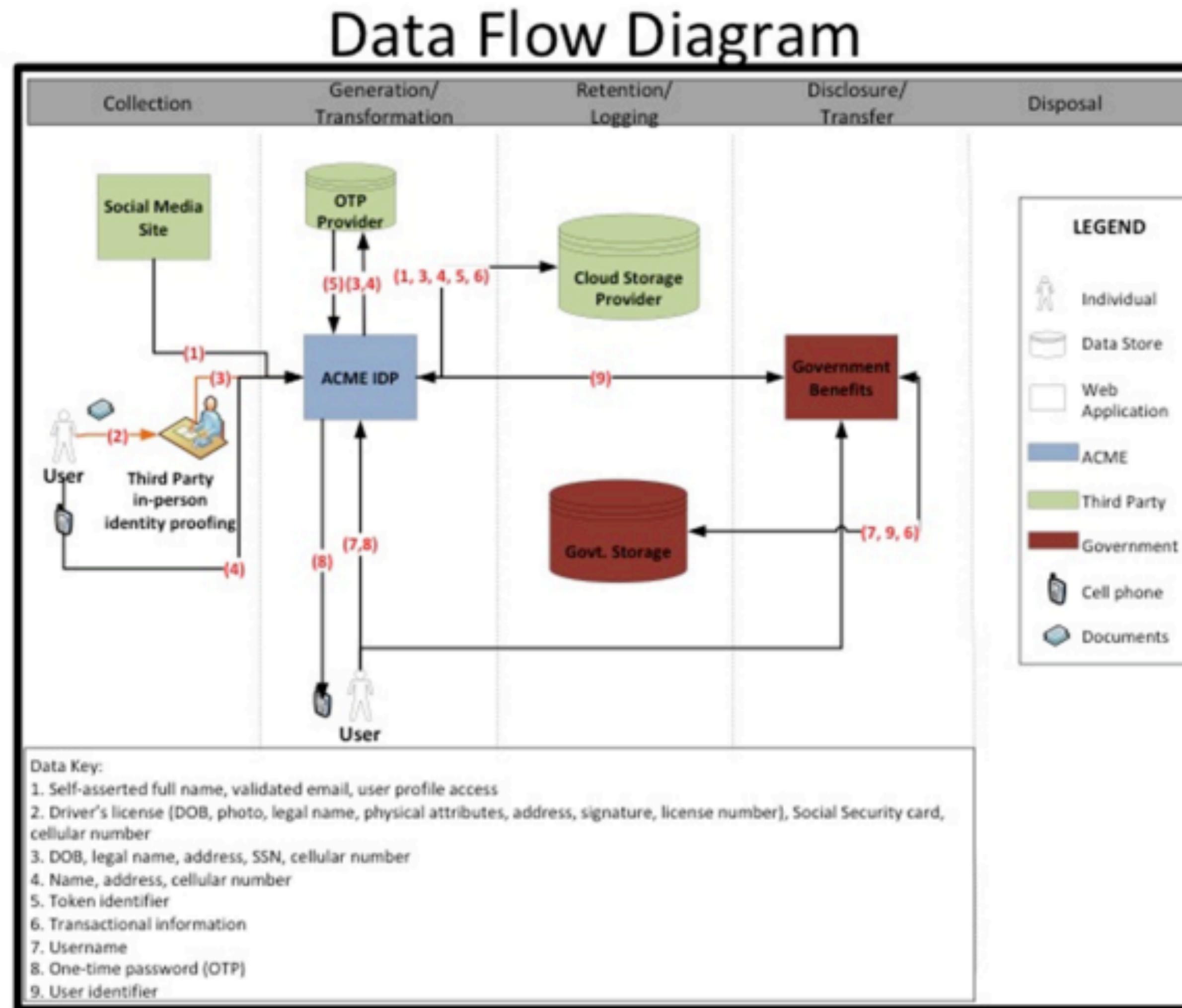


# PRAM - Task 1 - Map data processing within the system

## Use of credential to access benefits



# PRAM - Task 1 - Map data processing within the system



# PRAM - Task 2 - Catalog general contextual factors

Data Action	Personal Information	Specific Context	Summary Issues
Collection from the Social Media Site	<ul style="list-style-type: none"> <li>-Self-Asserted Full Name</li> <li>-Validated Email</li> <li>-List of Friends</li> <li>-Profile Photograph</li> </ul>	<ul style="list-style-type: none"> <li>-One-time action (per user) between social credential and ACME IDP, but establishes an ongoing relationship between user's social media presence and ACME IDP</li> <li>-Social credential linking is visible to user</li> <li>-Linking of social credential simplifies access to government benefits system</li> <li>-User profile may contain information the user considers sensitive</li> <li>-User profile may contain information from other users not participating in the system</li> <li>-User profile includes information unrelated to the purpose and operations of the system</li> <li>-Access to PI is consented by user</li> <li>-Nature of the API: full profile access is granted (by default: name, validated email, profile photograph, and list of friends)</li> </ul>	<ul style="list-style-type: none"> <li>-Full social credential profile access (including picture and list of friends) is not necessary for fulfilling operational purpose.</li> <li>-Will users understand the eventual high-assurance credential is controlled by ACME and not by their social credential provider?</li> <li>-How will perception of the social media organization's privacy practices impact users' willingness to consent to this data action?</li> <li>-Will the user understand ACME will have ongoing access to information stored in their social profile?</li> <li>-Will users' social media privacy settings allow this data action?</li> </ul>

# PRAM - Task 3 - Likelihood

Data Actions	Summary Issues	Problematic Data Actions	Potential Problems for Individuals	Likelihood
Collection from the social media site	Full social credential profile access (including picture and list of friends) is not necessary for fulfilling operational purpose.	-Appropriation -Induced disclosure -Surveillance -Unanticipated revelation	Stigmatization: Information is revealed about the individual that they would prefer not to disclose.  Power Imbalance: People must provide extensive information, giving the acquirer an unfair advantage.	7
	Will users understand the eventual high-assurance credential is controlled by ACME and not by their social credential provider?	-The summary issue will be associated with another data action.		2
	How will perception of the social media organization's privacy practices impact users' willingness to consent to this data action?	-Induced disclosure -Surveillance	Loss of Trust: Individuals lose trust in ACME due to a breach in expectations about the handling of personal information.	6

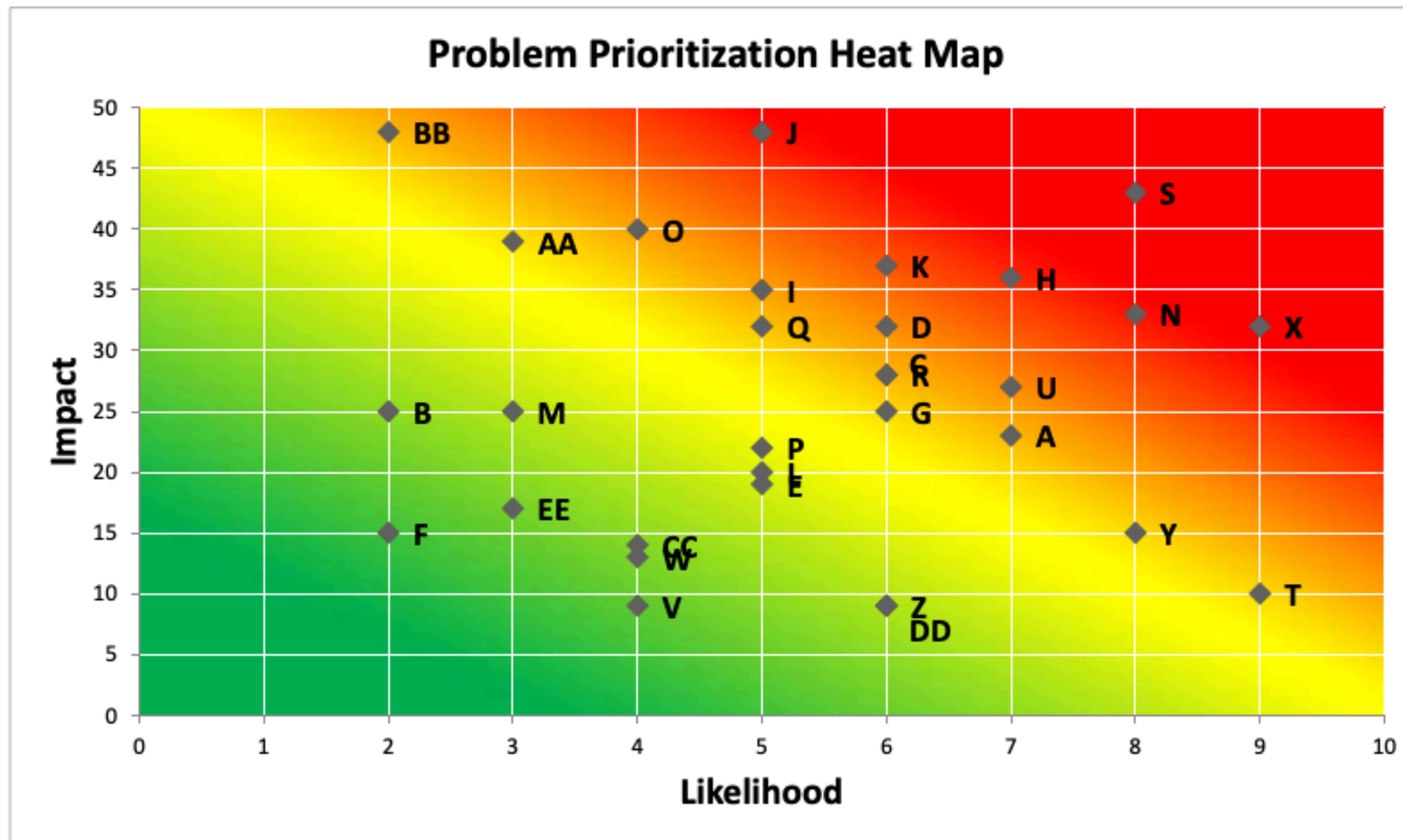
# PRAM - Task 4- Likelihood X Impact = Risk

Data Actions	Summary Issues	Problematic Data Actions	Potential Problems for Individuals	Business Impact Factors					Total Business Impact
				Noncompliance Costs	Direct Business Costs	Reputational Costs	Internal Culture Costs	Other	
Collection from the social media site	Full social credential profile access (including picture and list of friends) is not necessary for fulfilling operational purpose.	<ul style="list-style-type: none"> <li>-Appropriation</li> <li>-Induced disclosure</li> <li>-Surveillance</li> <li>-Unanticipated revelation</li> </ul>	Stigmatization	7	6	6	4		23
			Power Imbalance	7	6	8	4		25
	How will perception of the social media organization's privacy practices impact users' willingness to consent to this data action?	<ul style="list-style-type: none"> <li>-Induced disclosure</li> <li>-Surveillance</li> </ul>	Loss of Trust	7	6	8	7		28

# PRAM - Task 5 - Problem prioritization

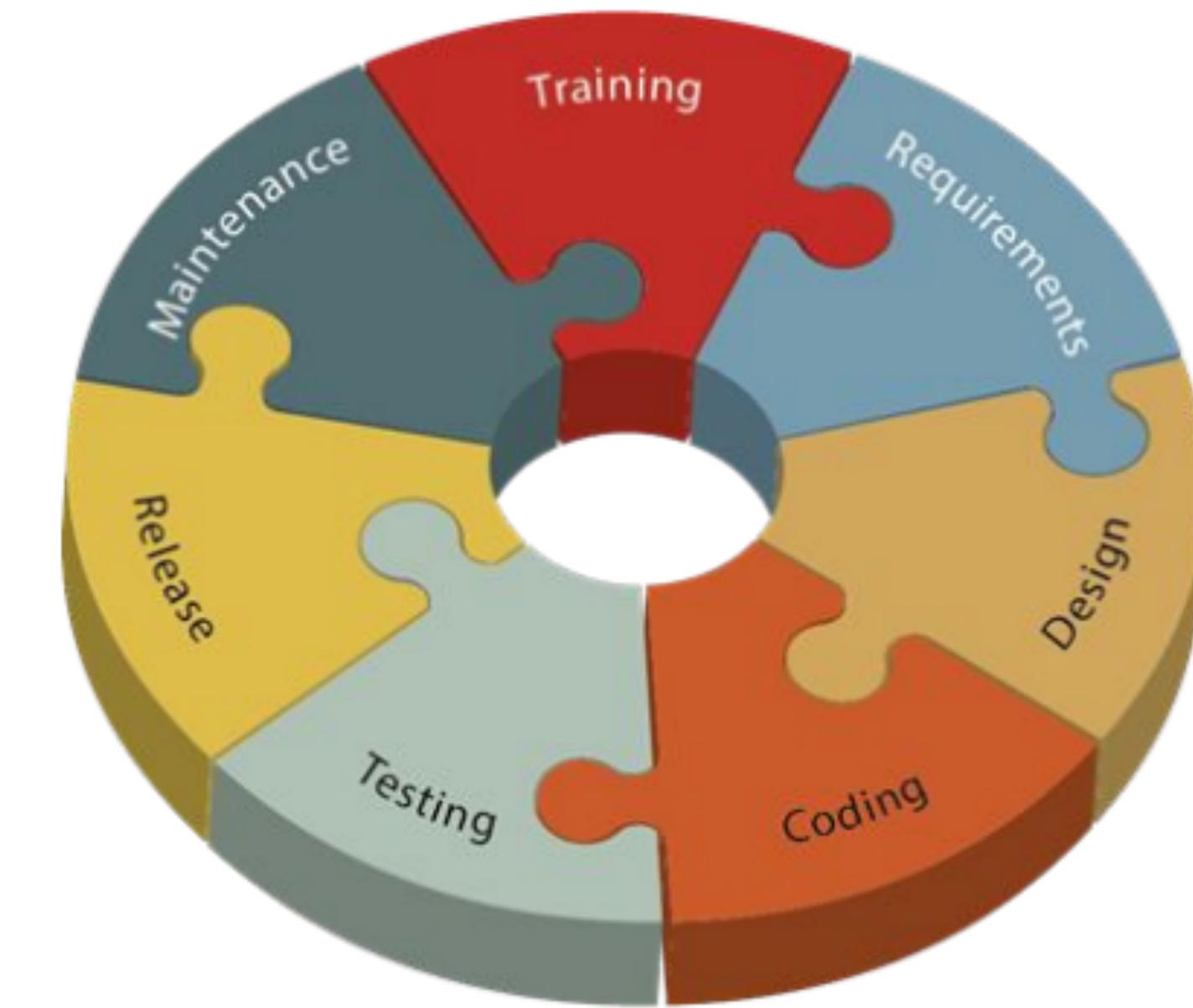
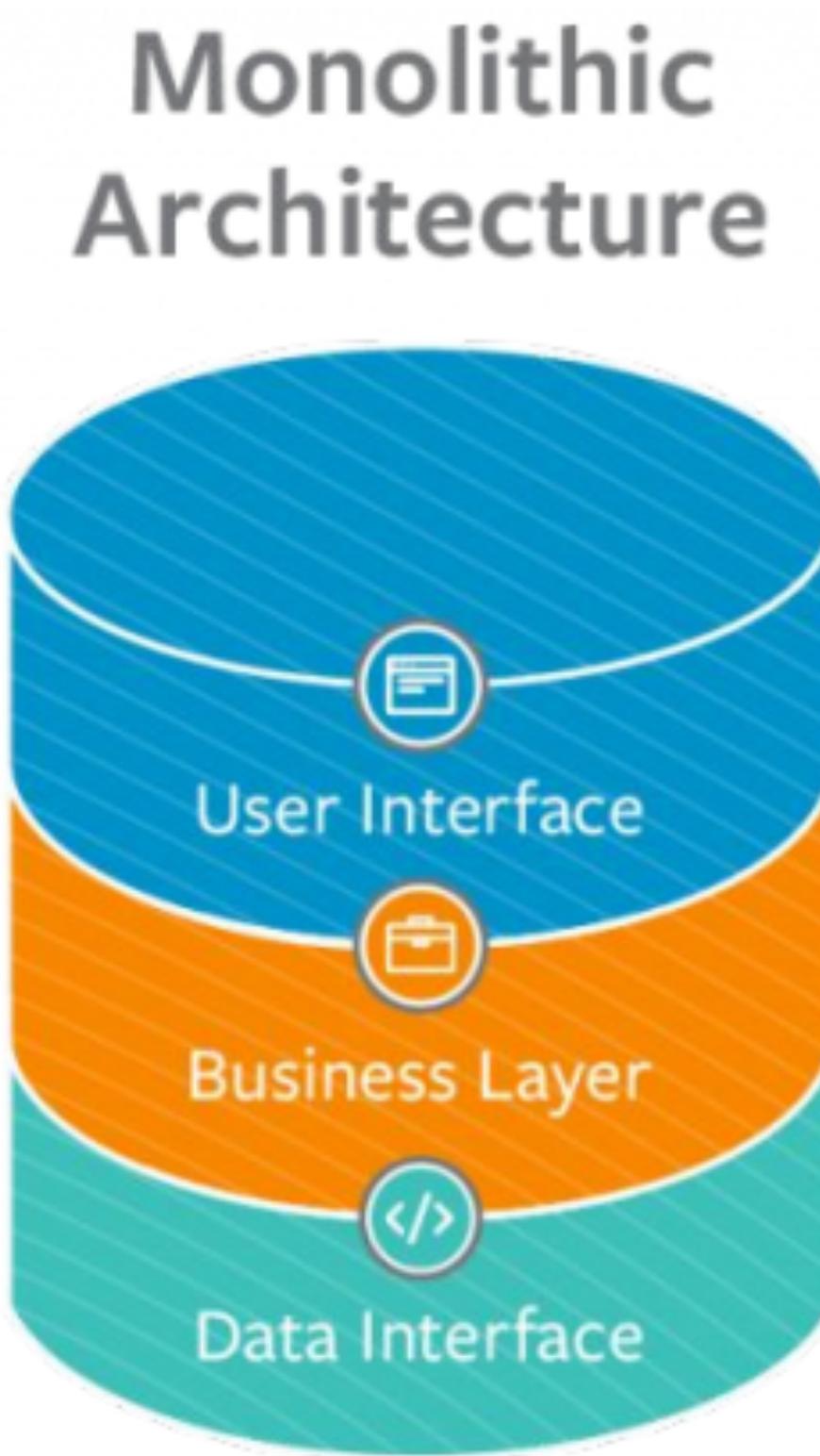
Data Actions	Potential Problems	Point Label	Likelihood	Business Impact
Collection from the social media site	Stigmatization	A	7	23
	Power Imbalance	B	2	25
	Loss of Trust	C	6	28
DA2	Economic Loss	D	6	32
	Loss of Autonomy	E	5	19
	Exclusion	F	2	15
DA3	Loss of Trust	G	6	25
	Stigmatization	H	7	36
	Loss of Liberty	I	5	35
DA4	Loss of Trust	J	5	48
DA5	Economic Loss	K	6	37
	Loss of Autonomy	L	5	20
	Power Imbalance	M	3	25
DA6	Exclusion	N	8	33
	Stigmatization	O	4	40
	Loss of Trust	P	5	22
DA7	Loss of autonomy	Q	5	32
	Exclusion	R	6	28
	Loss of Autonomy	S	8	43
DA8	Stigmatization	T	9	10
	Power Imbalance	U	7	27
	Exclusion	V	4	9
DA9	Loss of autonomy	W	4	13
	Stigmatization	X	9	32
	Power Imbalance	Y	8	15
DA10	Exclusion	Z	6	9
	Loss of Trust	AA	3	39
	Loss of Liberty	BB	2	48
DA10	Loss of Trust	CC	4	14
	Power Imbalance	DD	6	9
	Stigmatization	EE	3	17

# PRAM - Task 5 - Problem prioritization



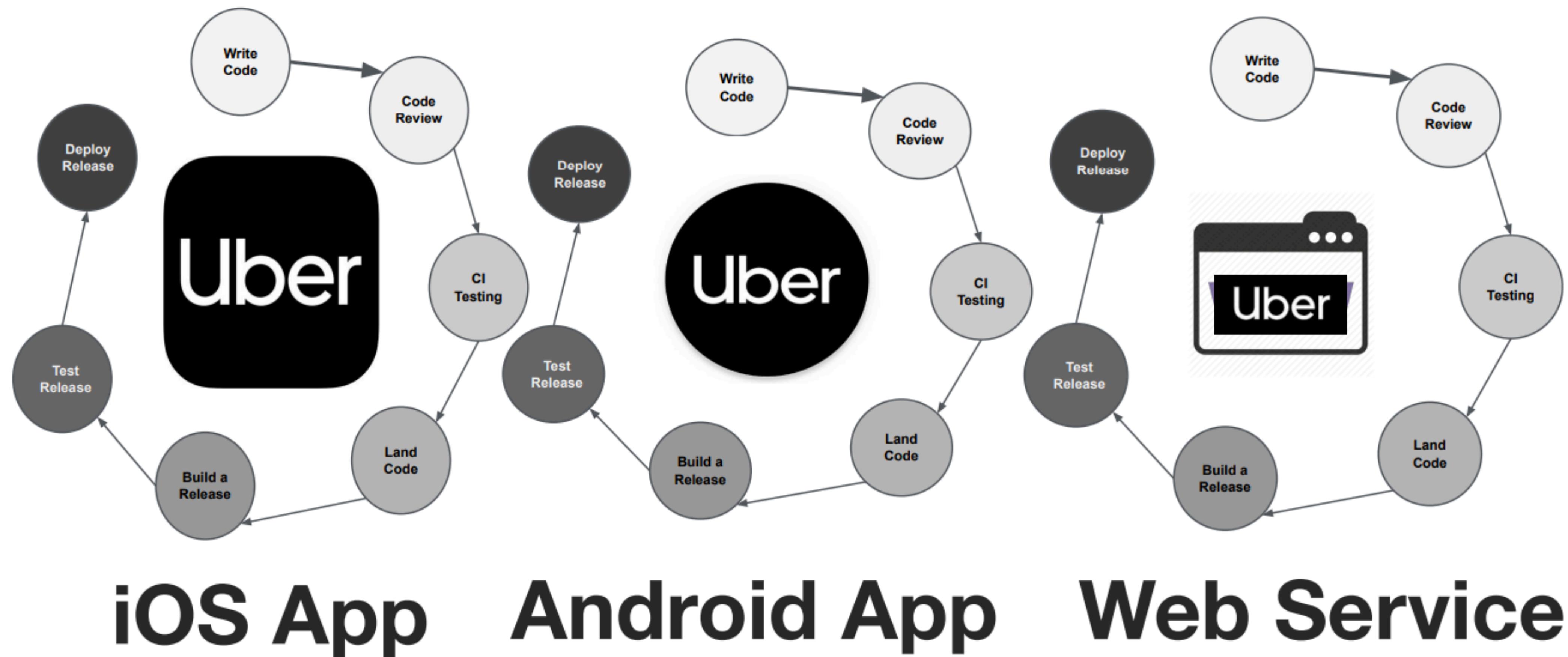
# Practices in the industry

# Existing Guidelines



Transparency	Accuracy	Data Minimization	Privacy Rights
Lawfulness	Purpose Limitation	Storage Limitation	Privacy by Default

# Microservices

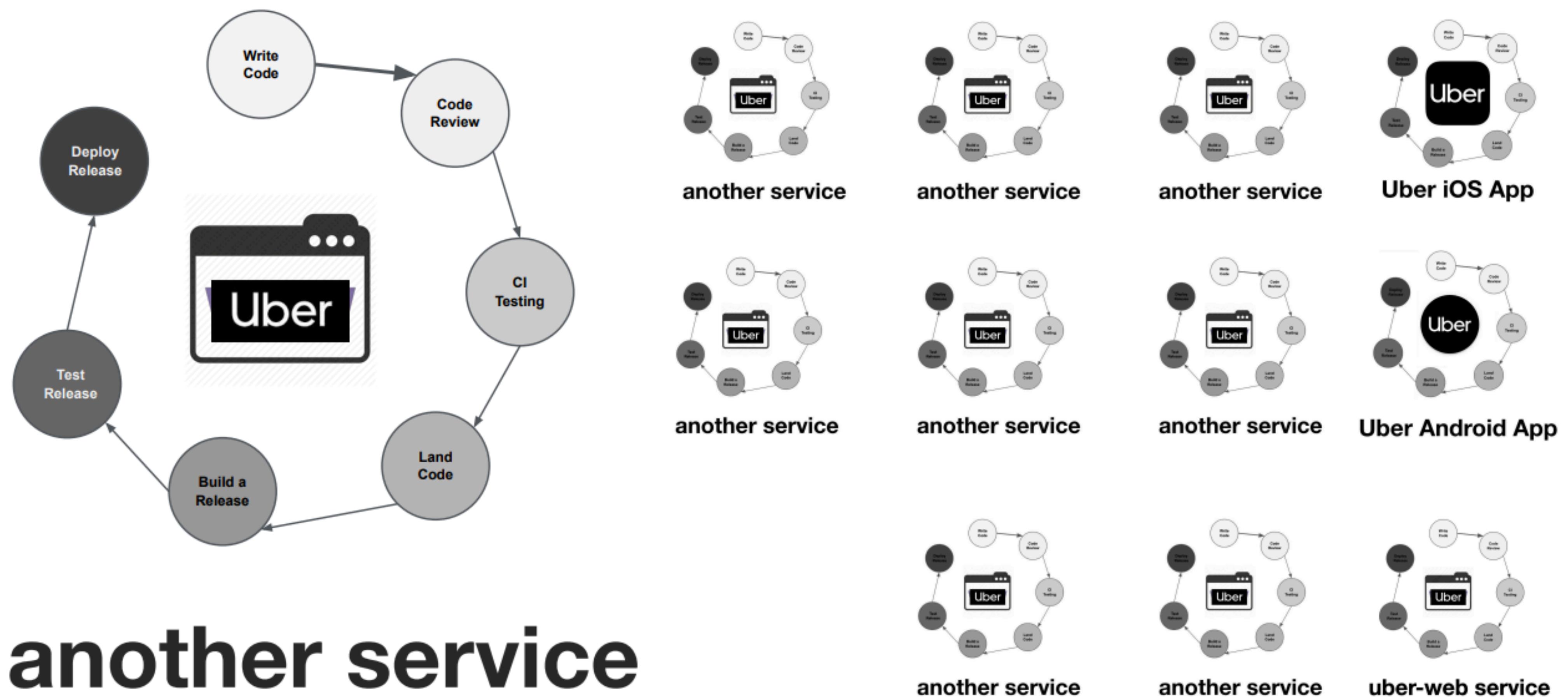


**iOS App**

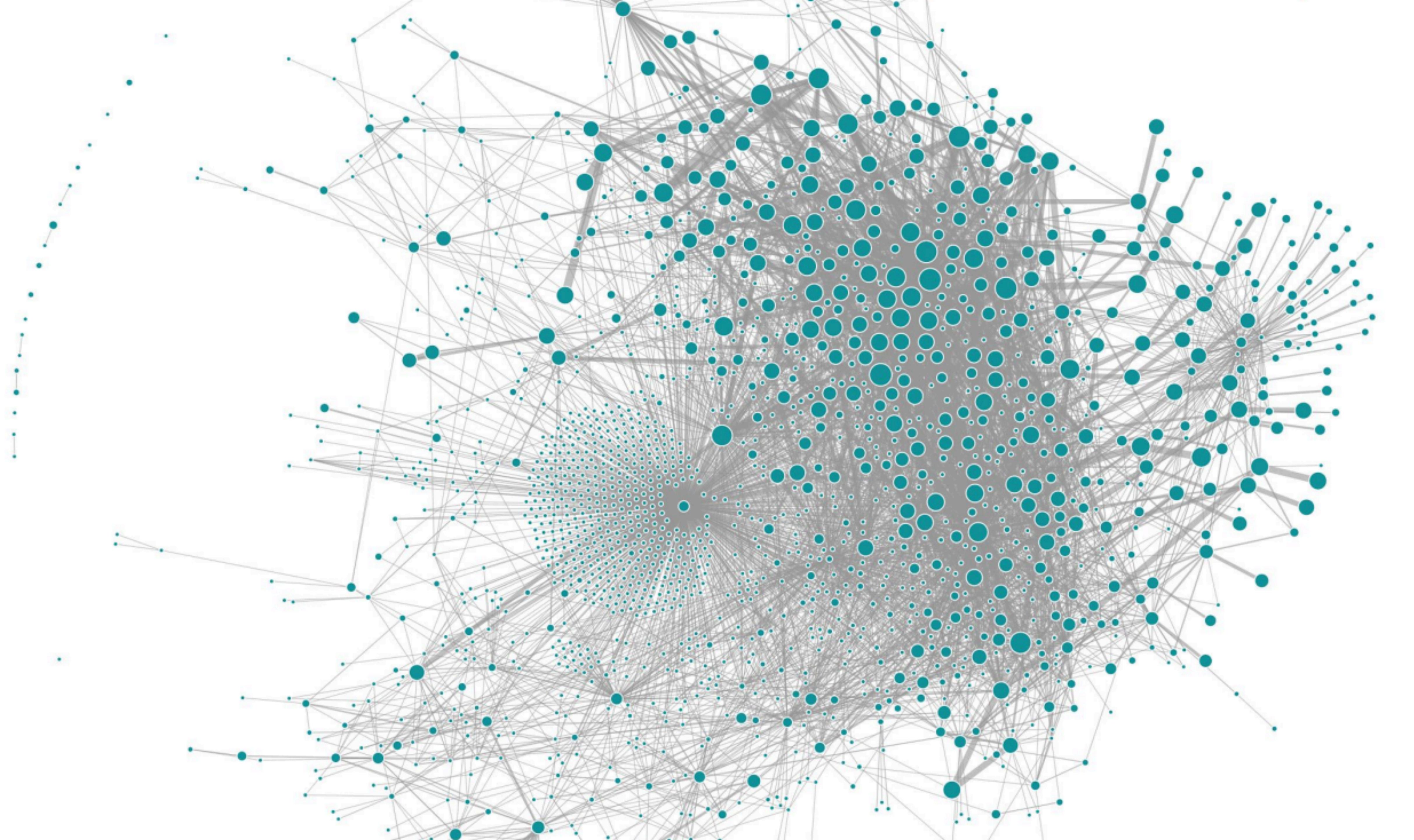
**Android App**

**Web Service**

# Many more services

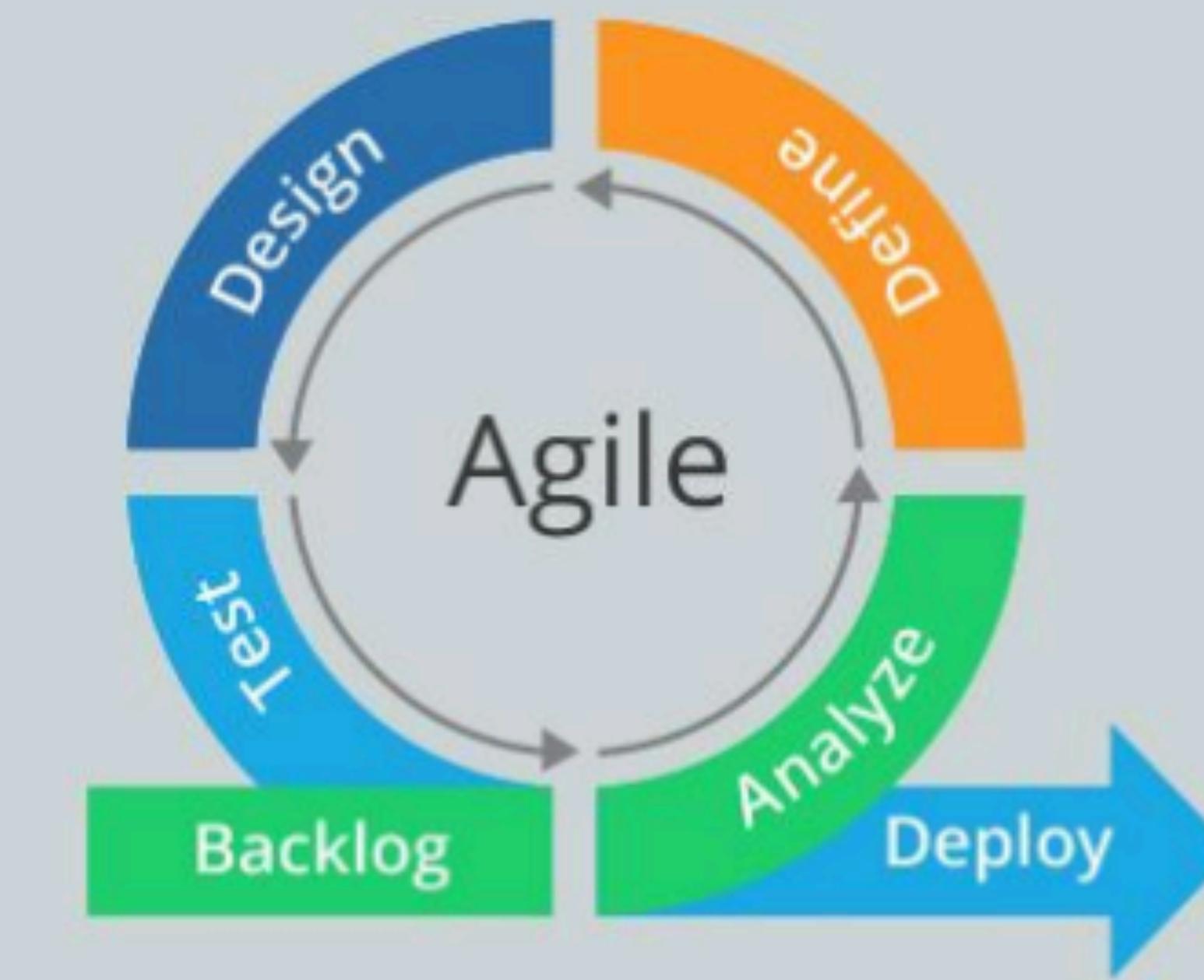


**another service**



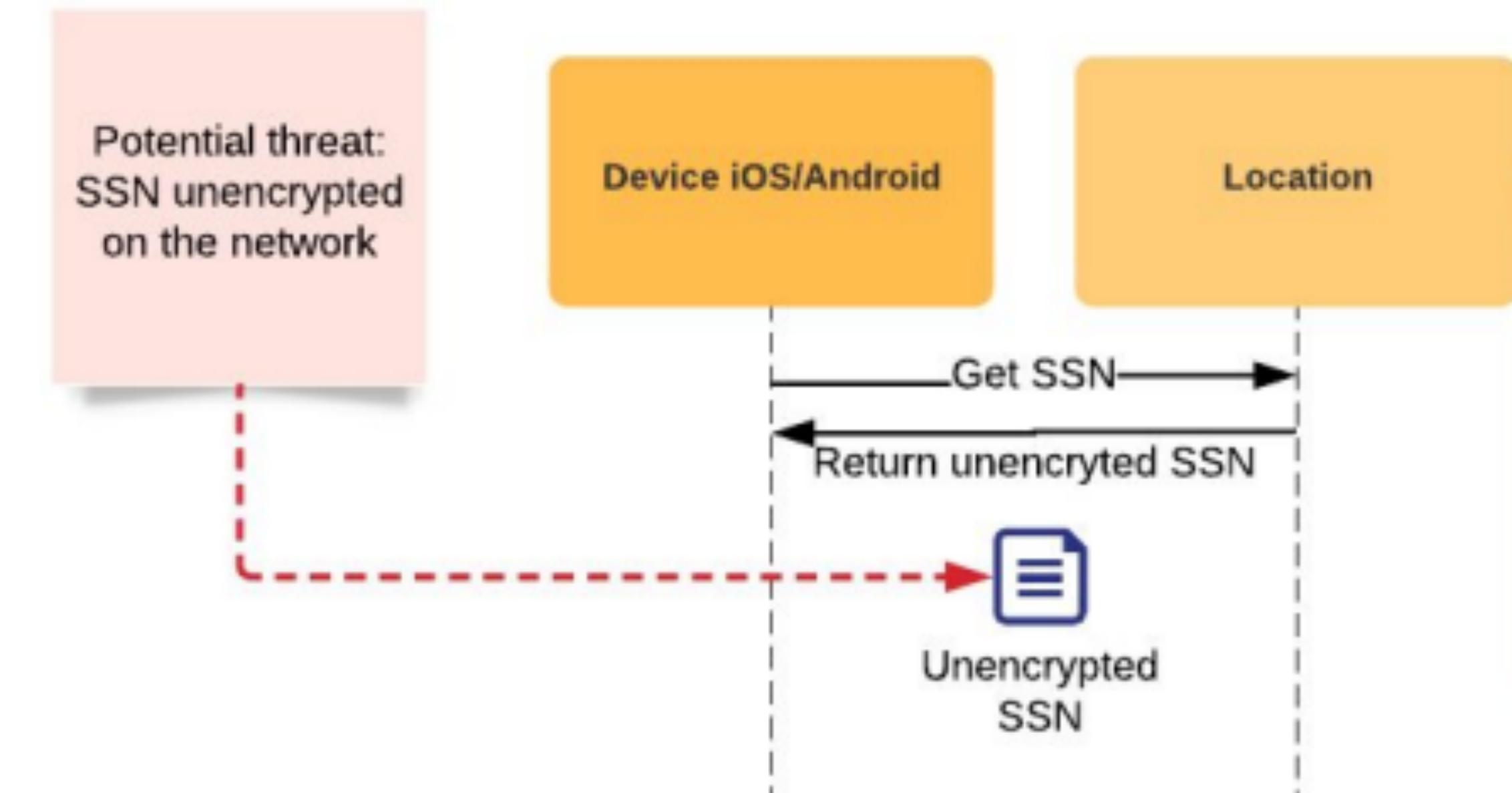
# Agile Development

## Waterfall vs. Agile



# Challenge 1: System characterization

- Distributed data
- Mix of structured + unstructured data
- Off-the-shelf tools do not scale
- No (stable) architectural documentation



# Data Classification

Tier 1: Highly Restricted

---

Tier 2: Restricted

---

Tier 3: Confidential

---

Tier 4: Public

# Example Category

Government Identifier & location

---

Vehicle Data

---

Non-Identifying Vehicle Data

---

Public Information

# Example Data Sets

Driver's License

---

License Plate Number  
Proof of Insurance

---

Make and Model  
Color

---

Product Brochures

# Stakeholders



# Data inventory

- Unified data category tags
- Automatic tagging and verification
- Maturity Levels: Tagging at DB level, tagging at column level, identify ALL data of an individual
- Use Data Inventory results to improve processes

# LegalEase

## Bootstrapping Privacy Compliance in Big Data Systems

Shayak Sen, [Saikat Guha](#), Anupam Dutta, [Sriram Rajamani](#), Janice Tsai, Jeannette Wing  
*Proceedings of the 35th IEEE Symposium on Security & Privacy (Oakland)* | May 2014  
Published by IEEE

 [Download BibTex](#)

With the rapid increase in cloud services collecting and using user data to offer personalized experiences, ensuring that these services comply with their privacy policies has become a business imperative for building user trust. However, most compliance efforts in industry today rely on manual review processes and audits designed to safeguard user data, and therefore are resource intensive and lack coverage. In this paper, we present our experience building and operating a system to automate privacy policy compliance checking in Bing. Central to the design of the system are (a) LEGALEASE—a language that allows specification of privacy policies that impose restrictions on how user data is handled; and (b) G ROK—a data inventory for Map-Reduce-like big data systems that tracks how user data flows among programs. G ROK maps code-level schema elements to datatypes in LEGALEASE, in essence, annotating existing programs with information flow types with minimal human input. Compliance checking is thus reduced to information flow analysis of big data systems. The system, bootstrapped by a small team, checks compliance daily of millions of lines of ever-changing source code written by several thousand developers.

[View Publication](#)

### Groups

[Mobility, Networks, and Systems](#)  
[Systems | India](#)  
[Cloud and Infrastructure Security Group](#)

### Research Areas

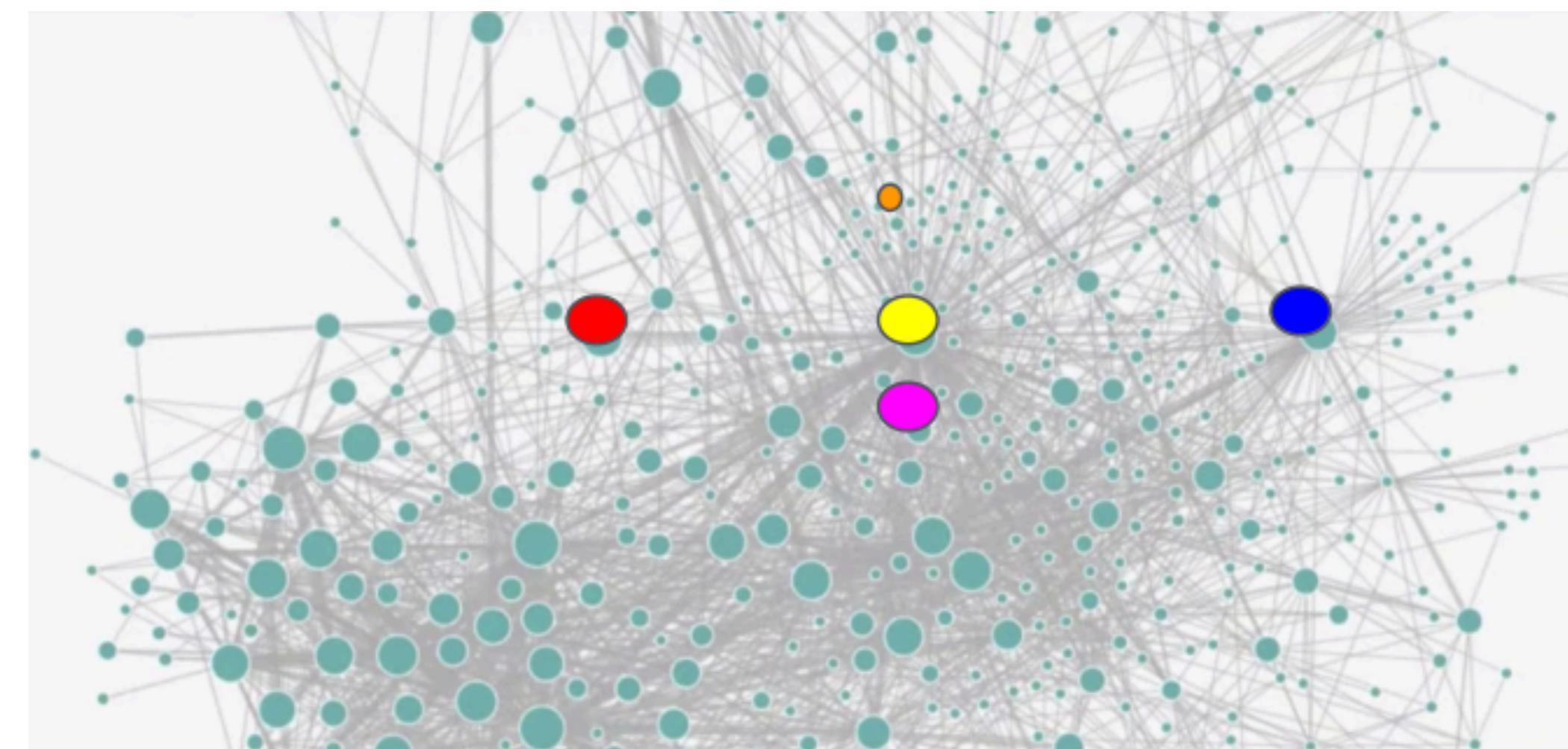
[Security, privacy, and cryptography](#)  
[Systems and networking](#)

### Research Labs

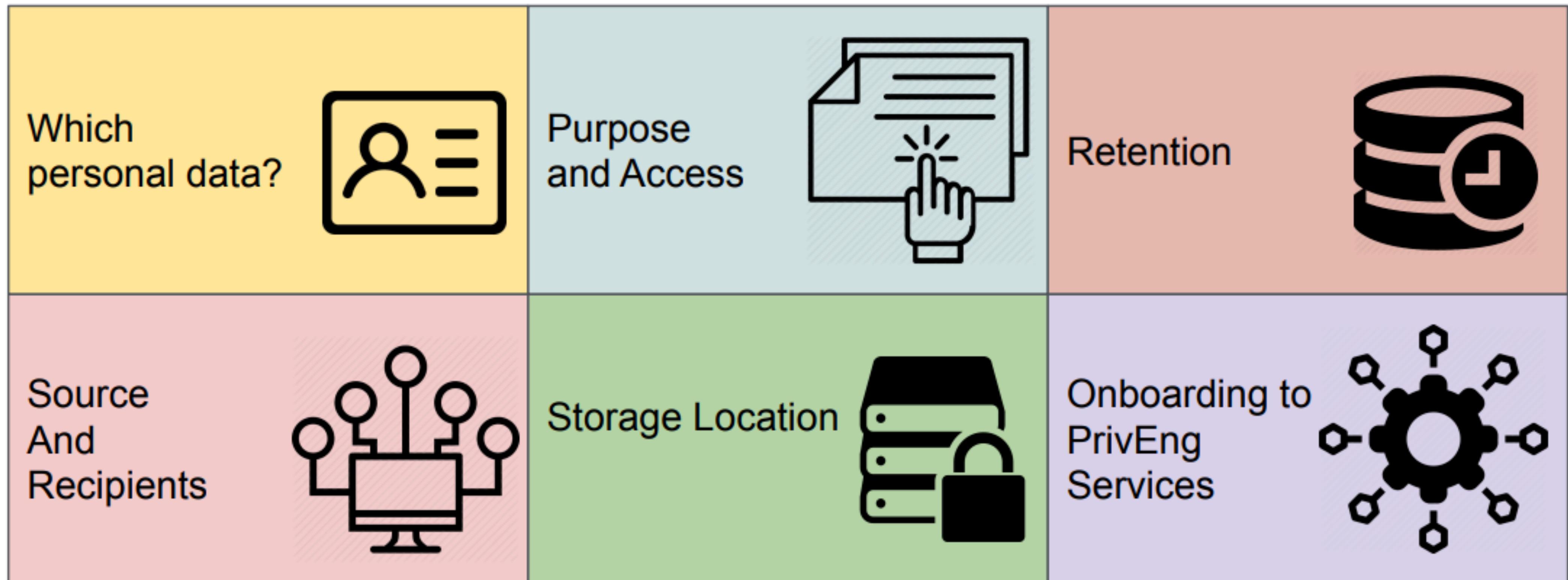
[Microsoft Research Lab - India](#)

# Challenge 2: Threats and Mitigation

- Privacy threat of the service vs the whole chain
- Where to place the control?
- Privacy can be slow vs Agile fast
- Legacy systems and privacy debt
- Resource and costs



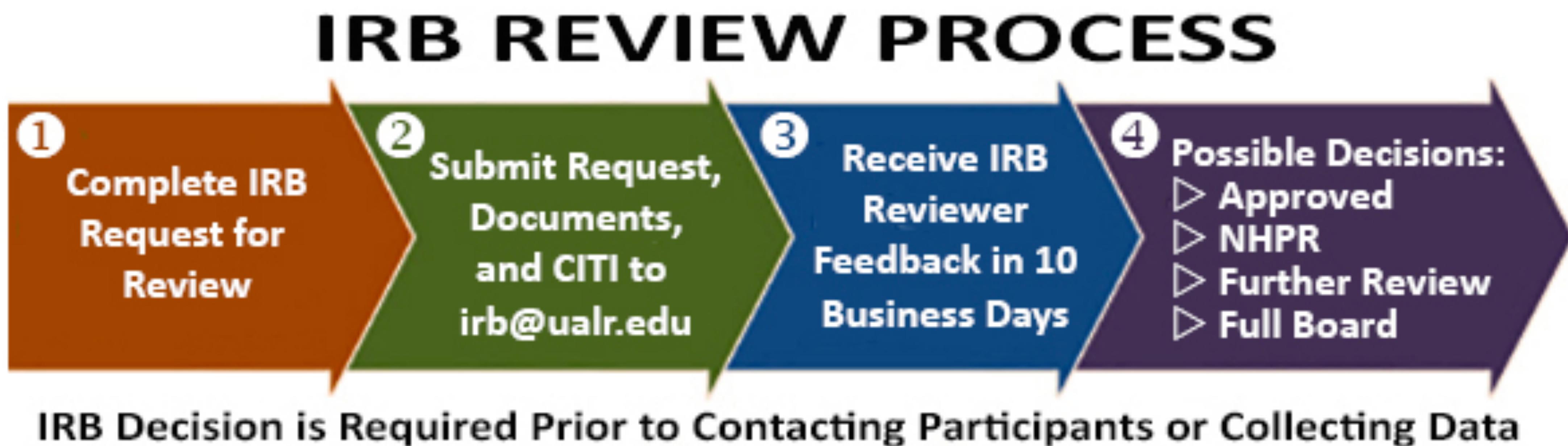
# Modular Reviews: Technical Privacy Consulting



# Modular Reviews: Technical Privacy Consulting

- Outcome:
  - Technical privacy requirements for this specific project
  - Mitigation prioritization
  - Input to Privacy Legal
- After the review:
  - Further analysis of platforms based on discovered knowledge
  - Embed privacy into platforms
  - Update Data Classification and Handling Standard

# IRB Review



[CLICK TO EXPAND](#)

# IRB Review



**Privacy Office****Authorities and  
Responsibilities**[Contacts](#)

## Authorities and Responsibilities of the Chief Privacy Officer

The activities of the Privacy Office serve to build privacy into departmental programs. The following is a framework of privacy laws through which the Privacy Office accomplishes its activities and mission:

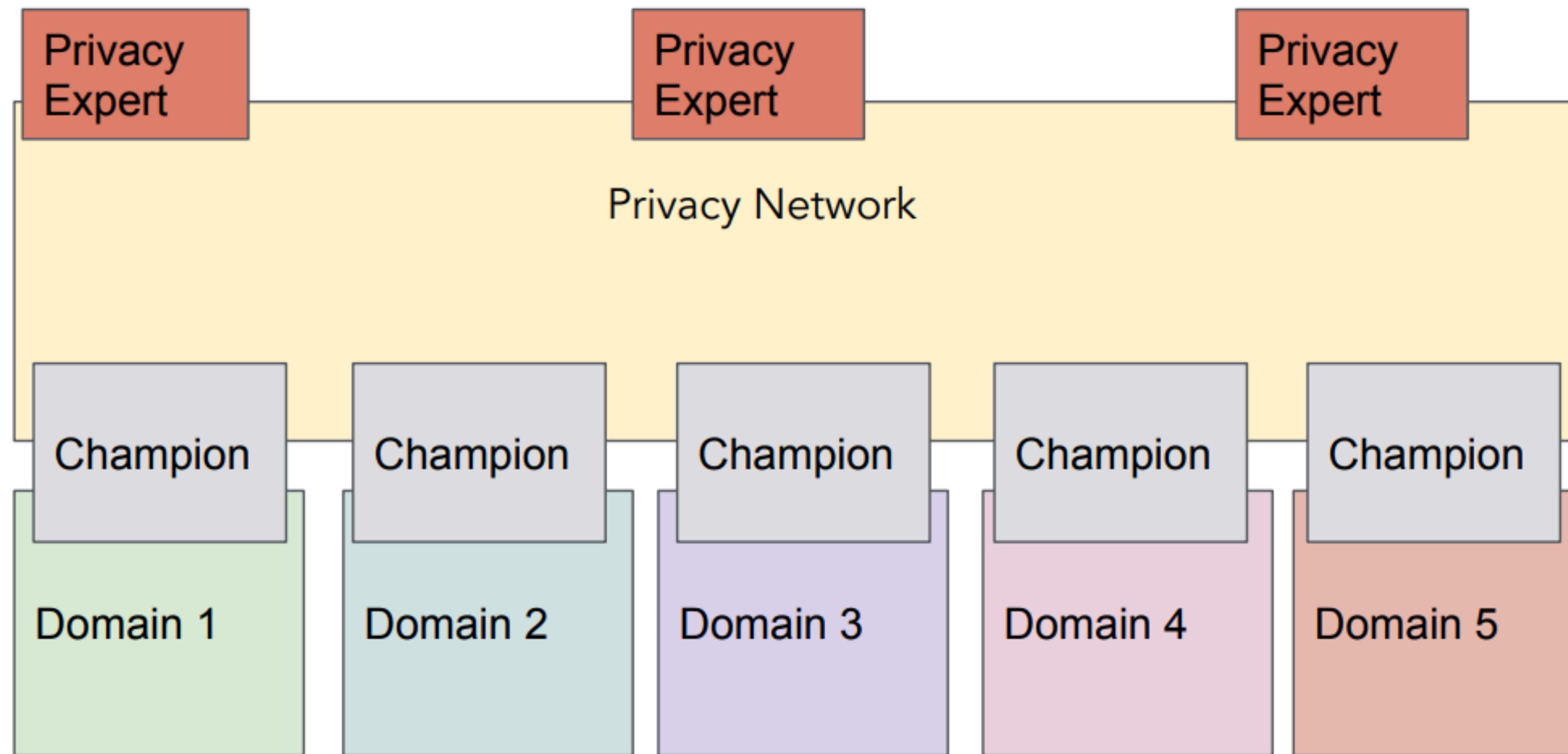
- Privacy Act of 1974, as amended (5 U.S.C. § 552a): Embodies a code of fair information principles that governs the collection, maintenance, use, and dissemination of personally identifiable information by federal agencies;
- E-government Act of 2002 (Public Law 107-347): Mandates Privacy Impact Assessments (PIAs) for all Federal agencies when there are new collections of, or new technologies applied to, personally identifiable information;

---

[Home](#) > [Security](#)**SECURITY ADVISER**By [Roger A. Grimes](#) | [Follow](#)

## Why your company needs a chief privacy officer

# Uber Approach: Education and Privacy Champions



# Chief Privacy Officers

- Companies are increasingly appointing CPOs to have a central point of contact for privacy concerns
- Role of CPO varies in each company
  - Draft privacy policy
  - Respond to customer concerns
  - Educate employees about company privacy policy
  - Review new products and services for compliance with privacy policy
  - Develop new initiatives to keep company out front on privacy issue
  - Monitor pending privacy legislation

# Motivations: Champions



- “Champions” advocate for a cause (e.g., an innovation or idea), encourage others to engage, and aid with overcoming barriers that a new idea could face
- Literature in software engineering shows champions’ value in promoting both security and new software technologies
- And what about privacy?

# Privacy Champions



- Formally or informally promote best practices for users' privacy, educate others, persuade, and advocate for privacy adoption throughout the software development process
- Have an official or unofficial role within their team acting as the “voice” of users' privacy for the product or team, for example by giving privacy-related advice that can influence decisions and privacy practices

# Common Barriers for Implementing Privacy in Software Design

- Negative privacy culture and attitudes (e.g., “I’ve got nothing to hide”)
- Tensions between privacy and business priorities
- Lack of standardisation, evaluation metrics, and automated privacy tools
- Technical complexity

# Strategies for Promoting Privacy

## Effective

- ❖ Regular privacy-focused meetings and informal discussions
- ❖ Management support, facilitation of communication among stakeholders (e.g., between legal and product teams)
- ❖ Appropriate privacy documentation and guidelines
- ❖ Incorporating privacy considerations into design reviews

## Not effective

- ❖ Punishing developers for not implementing privacy features
- ❖ Company-wide awareness programs or on-boarding privacy training for new hires

# How to Support and Attract Privacy Champions

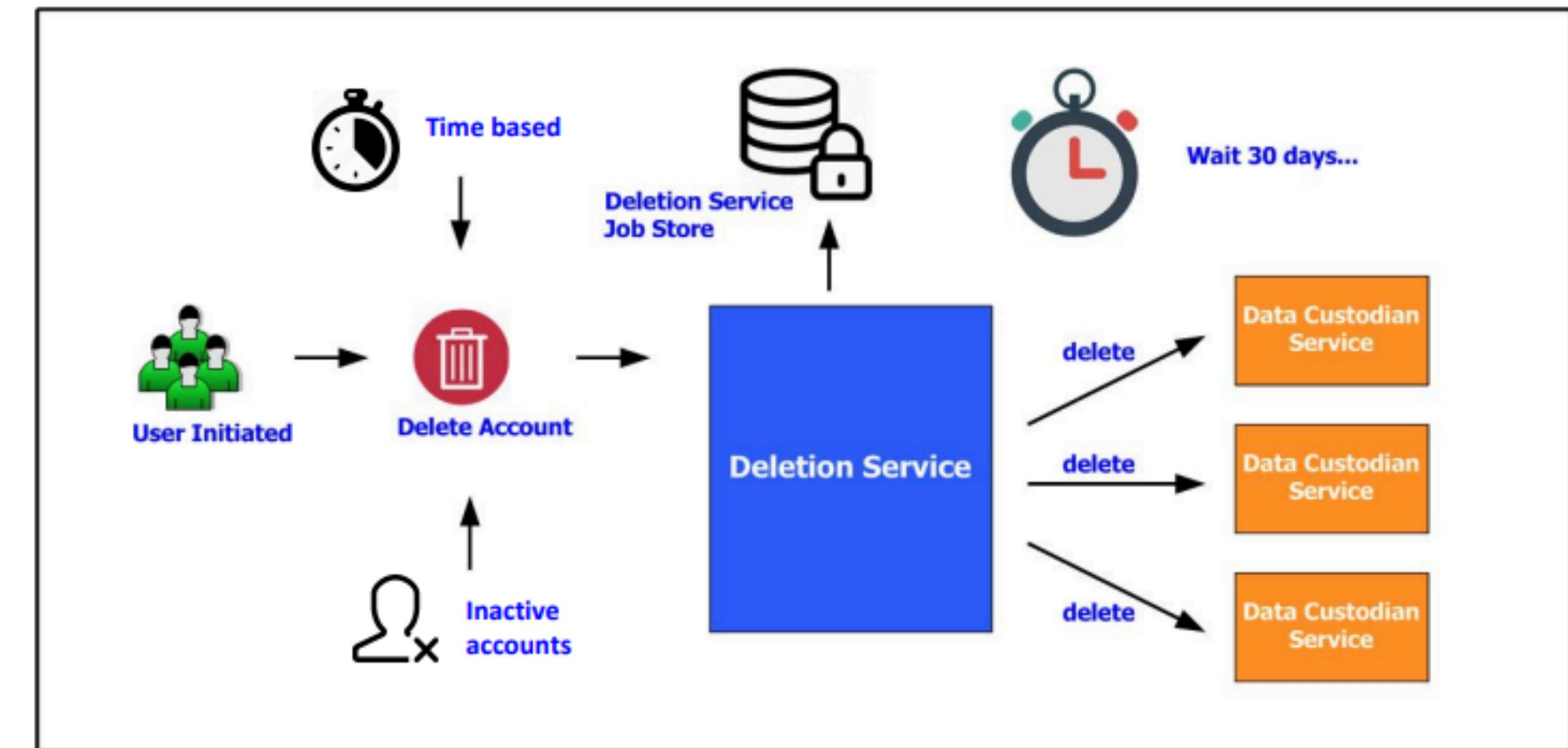
- ❖ Embed privacy values in the organisational culture
- ❖ Include privacy topics in university degree and online learning curricula
- ❖ Offer hands-on privacy-oriented projects (e.g., privacy hackathons)
- ❖ Acknowledge privacy-oriented efforts (by both colleagues and management)
- ❖ Provide resources (compensate extra time, or let spend 10-20% of their time on privacy work)

# Challenge 3: Doing Privacy at Scale: Deletion

- Multiple use cases: user initiated account deletion, inactive account deletion, time-based deletion
- Variety of data stores
- Scalable, reliable, adaptable, demonstrable

# Uber's Approach to Data Deletion

- Support scale of data, data stores, and microservices
- Privacy Impact Assessment and Technical Privacy Reviews
- Vetting process combines legal and technical privacy
- Automate onboarding process for new services



# Data governance

- People, process, and technology for managing data within an organization
- Data-centric threat modeling and risk assessment
- Protect data throughout information lifecycle – Including data destruction at end of lifecycle
- Assign responsibility

# Recap

- Why
- Privacy by Design
- Practices in the industry
- Privacy Reviews

# Credits

1. PRIVACY AT SPEED: PRIVACY BY DESIGN FOR AGILE DEVELOPMENT AT UBER,  
<https://www.usenix.org/conference/enigma2020/presentation/bozdag>
2. Privacy Policy, Law, and Technology, Lorrie Cranor, <https://cups.cs.cmu.edu/courses/pplt-fa15/>