

Privacy by Design in the SDLC

Why, When, How?



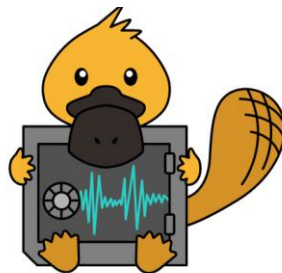
PRESENTER

Cat Easdon

Senior Privacy Engineer

whoami

- Senior Privacy Engineer and TechEvangelist at Dynatrace
- 🏔️-obsessed Brit based in Innsbruck
- Outside of work: research and tech policy, trail running, hiking, skiing...
- Previously: hacking CPUs at TU Graz



Outline

Why

- What does privacy mean?
- Privacy threats and harms
- What even is personal data?

How

- The 7 Privacy by Design (PbD) principles
- Practical examples at the code-, API-, and UI-level

When

- Overview of PbD in the software development lifecycle



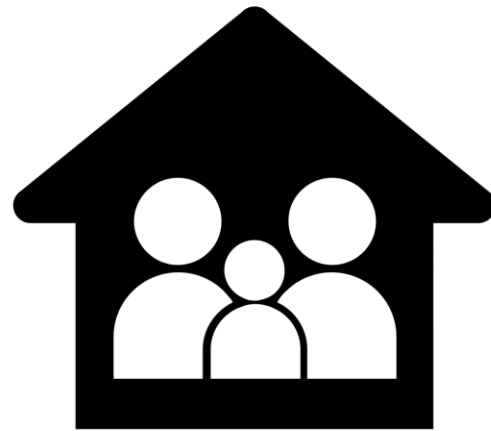
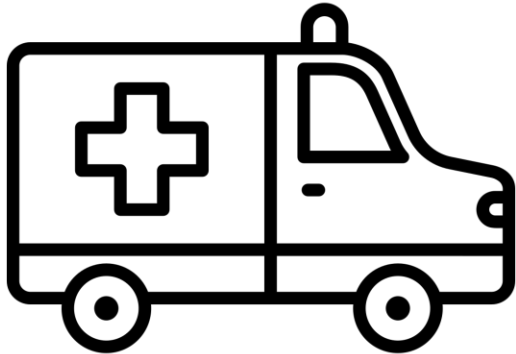
Copyright Notice

- **Dynatrace content and branding:** © 2024 Dynatrace LLC
- **Third-party images:** see links for attribution
- **Unattributed images:** generated with [DALL·E 3](#) or used under license from the [Noun Project](#)
- **All other content:** original work by the author, may be reused with attribution

Why?

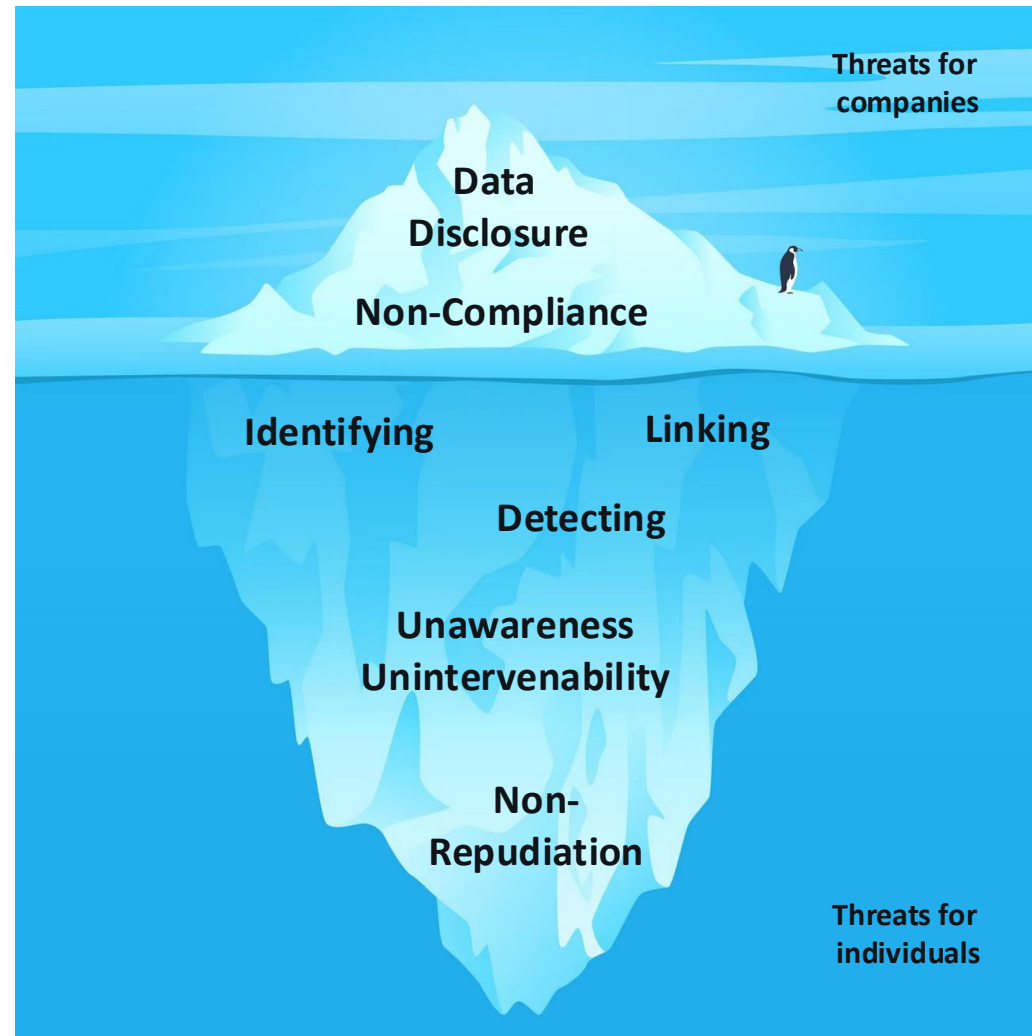
What does privacy mean?

- Privacy isn't about hiding everything!
- Society runs on information flows, but these flows should be *context-appropriate*
- Data type, data subject, sender, recipient, legal basis, and purpose all matter



Privacy threats: LINDDUN

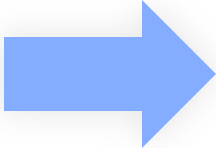
LINDDUN



Privacy harms: manipulation

Neuroticism - Trapped Neuroticism - Stress Reactors Neuroticism - Self Lovers
Neuroticism - Easily Deflated Neuroticism - Internal Escapists
General Attitudes - I generally get a raw deal out of life Dealing with Stress - Hot and Cold
Dealing with Stress - Emotional Dealing with Stress - Bottled Up

Clickagy > Health > **Addictions** > Alcohol



Skydeo > ConditionGraph > Disease Propensity by Type > **Depression** Diagnosis

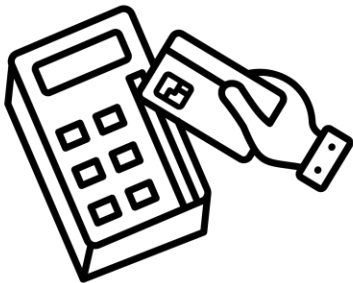
adready_drug_rehab (Grapeshot)

VisualDNA Lifestyle - Lifestyle - Health - Trying to cut down on **Alcohol**
Data provider: Nielsen Marketing Cloud



TransUnion - Demographics - Marital Status - Likely Recently **Divorced**

Eyeota - US Acxiom - CPG - **Alcoholic** Drinks - Vodka Brand - Grey Goose for age 21+ - Likely



What even is personal data?

- Name
- **UUIDs and any other unique identifiers**
- {Email, IP} address
- Location
- Education or employment details
- Inferences about a person (profiling)

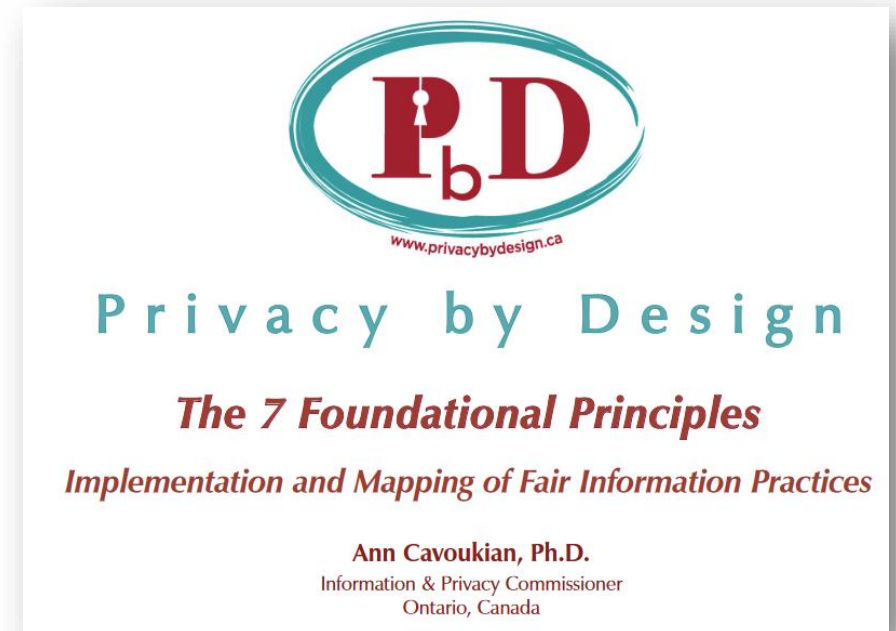
Sensitive personal data:

- Ethnic origin
- Citizenship or immigration status
- Religion, political views, beliefs
- Genetic and biometric data
- Health data
- ...

How?

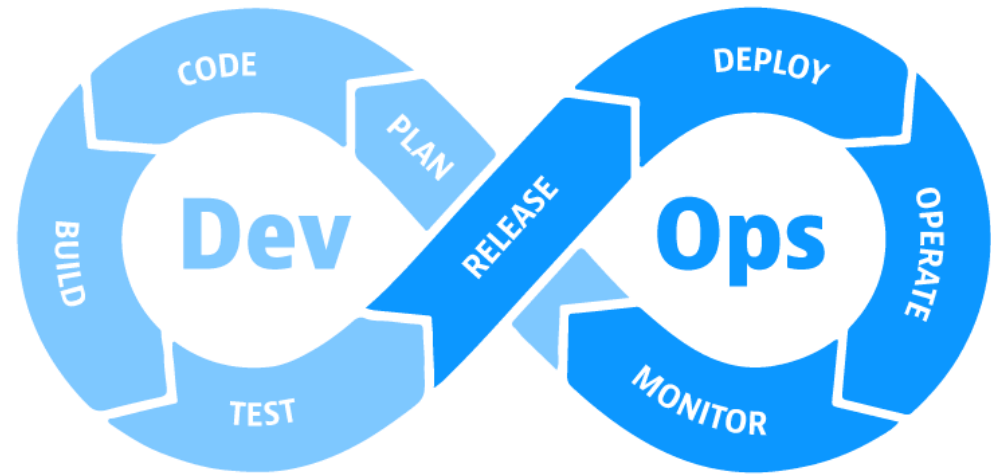
Privacy by design

1. Proactive not reactive; preventative not remedial
2. Privacy as the default
3. Privacy embedded into design
4. Full functionality – positive-sum, not zero-sum
5. End-to-end security – full lifecycle protection
6. Visibility and transparency
7. Respect for user privacy



1. Proactive not reactive; preventative not remedial

- Focus on proactively protecting users' privacy, instead of offering remedies when things go wrong
- Privacy controls built into the development lifecycle
- Commitment from org leadership and stakeholders
- Culture of continuous improvement



2. Privacy as the default

- If the user does nothing or takes the default action, then their privacy should be preserved
- Sharing data should always be an informed choice by the user
- Process the minimum amount of personal data necessary in a lawful, secure, and transparent way, and delete data once no longer needed



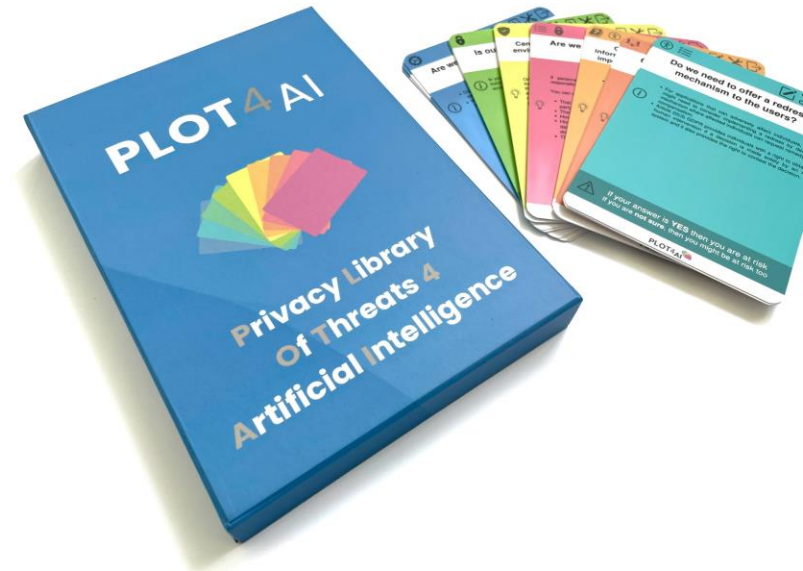
Case study: public by default

- New architectural requirement for our Document Service: documents should be **public by default**
- Let's update the DTO!
- What could go wrong here?

	@NotBlank
-	@JsonProperty(value = "isPublic")
-	boolean isPublic,
+	@JsonProperty(value = "isPrivate")
+	boolean isPrivate,

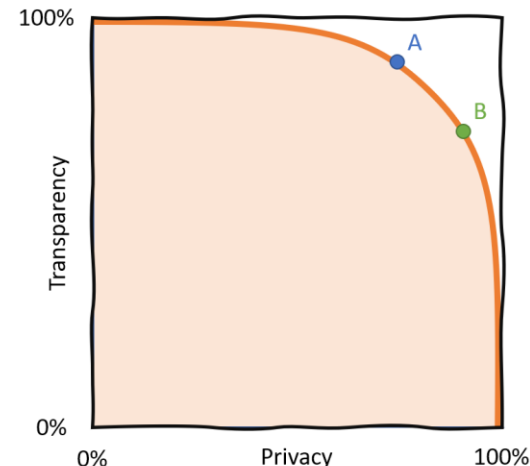
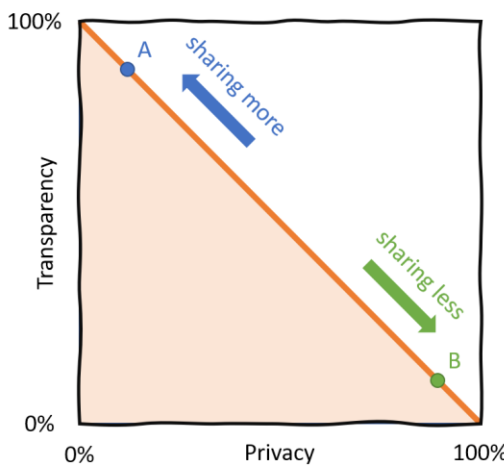
3. Privacy embedded into design

- Proactive + preventative *from the earliest design stages*
- Privacy threat modeling considering the product's broader context
- Minimize privacy impacts and demonstrate this by publishing impact assessments
- Creative solutions where requirements conflict



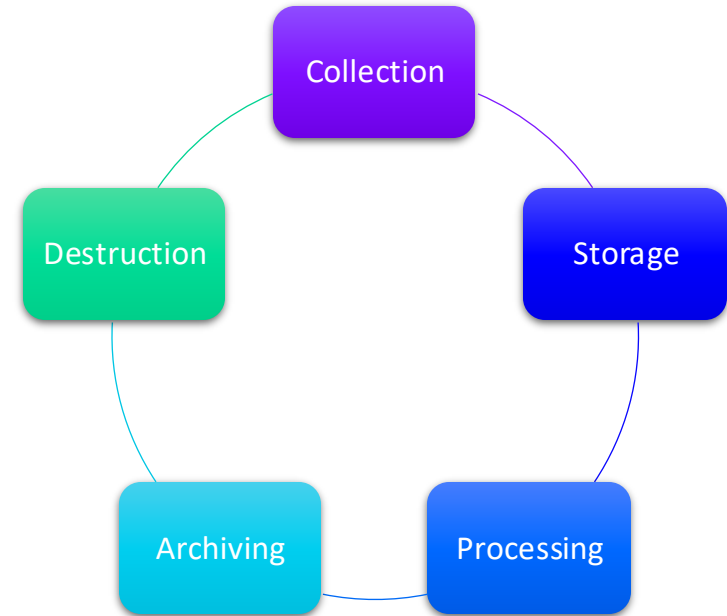
4. Full functionality – positive-sum, not zero-sum

- Remember those creative solutions earlier?
- Try to move beyond security vs. privacy or profit vs. privacy
- What would a win-win design look like?
- PETs play an important role here
 - Recommended: [Our Privacy Opportunity](#) by OpenMined



5. End-to-end security – full lifecycle protection

- Security controls throughout both the software development lifecycle *and the data lifecycle*
- Storage limitation: personal data must be securely deleted when it is no longer “needed” (keep purpose limitation in mind!)
- Common blind spot: backups and disaster recovery



Common issues in the data lifecycle

```
final var response = startQueryForUser(user, tenantInfo).block();
final LocalDateTime retrievedAt = LocalDateTime.now(ZoneOffset.UTC);
if (response == null) {
    log.error("Request to tenant {} for user {} failed!",
        tenantInfo.url(), user.email());
    return new ResultAtTime<>(retrievedAt, result: null);
} else if (response.requestToken() != null) {
    // ...
}
```

```
{
  "object": "user",
  "id": "uuid-redacted",
  "email": "a.sample.email@email.com",
  "name": "Cat Easdon",
  "picture": "https://lh3.googleusercontent.com/a/url-redacted",
  "created": 1702548524,
  "phone_number": "+431234567890",
  "groups": {"object": "list"...}
}
```


Common issues in the data lifecycle

Apply **purpose limitation** in API and interface design. Most systems don't need access to directly identifiable personal data and can work with pseudonyms (like UUIDs) instead:

```
sendEmail(userUuid, content)
```

vs.

```
sendEmail(userEmail, content)
```

Common issues in the data lifecycle

Where it is necessary to have direct access to personal data, ensure that it is not included in endpoint paths:

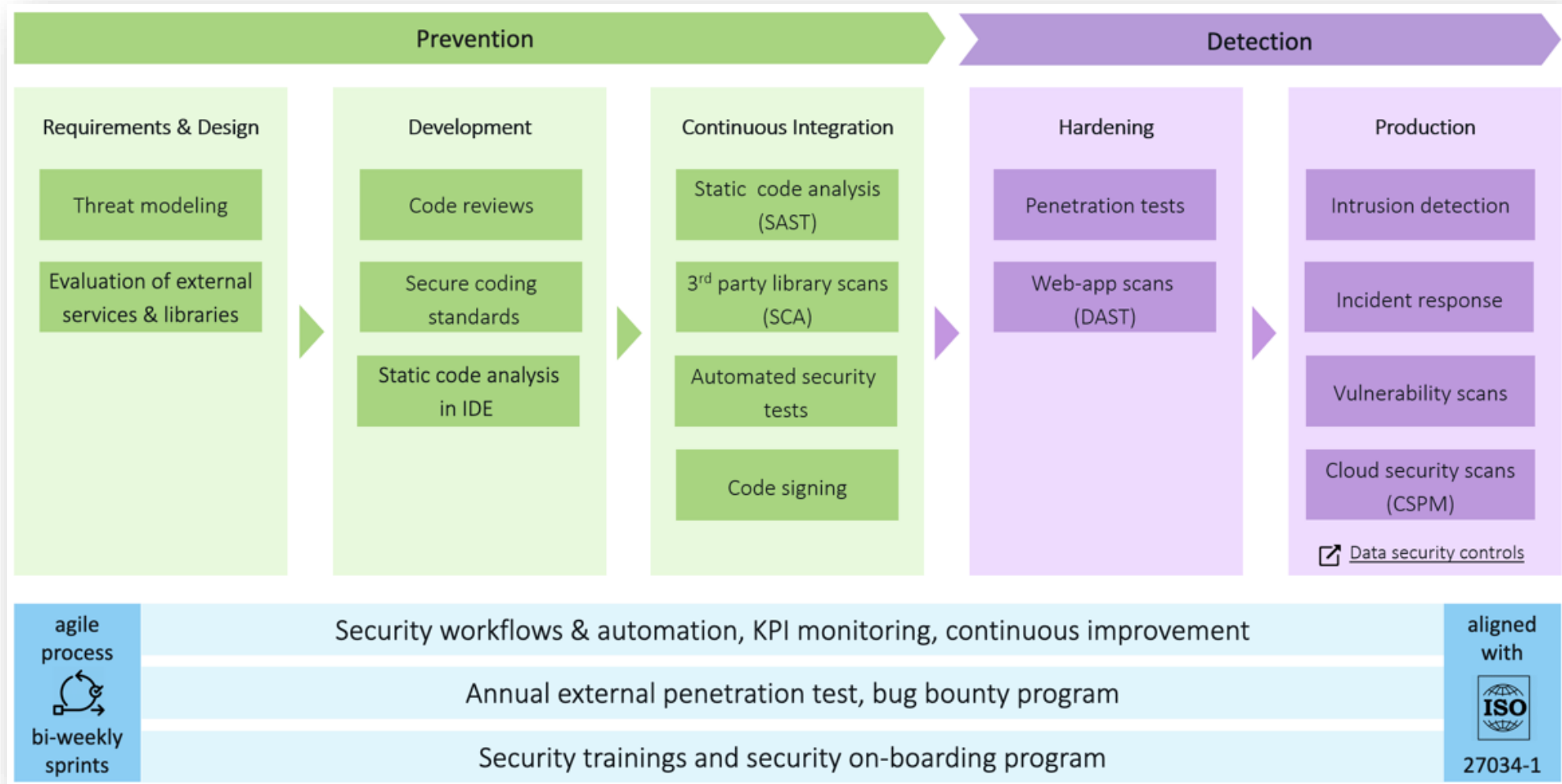
```
privacy-rocks.com/settings/maria@email.com
```

```
privacy-rocks.com/confirmation?user=andreas@email.com&token=ab456d
```

Otherwise, this data may be stored in the browser history and server logs, and gets sent to:

- Proxies and CDNs
- Google (or other ad vendor) and analytics providers
- Observability tooling
- ...

Security controls in the SDLC



Privacy as code

How to keep track of personal data?

- Internal policies
- Code-level annotations
- Data and software catalogues
- Personal data detection and data-flow mapping



```
public @interface MicrophoneAnnotation {  
    String ID();  
    Visibility[] visibility();  
    MicrophoneDataType[] dataType();  
    MicrophonePurpose[] purpose();  
    String [] purposeDescription();  
}
```

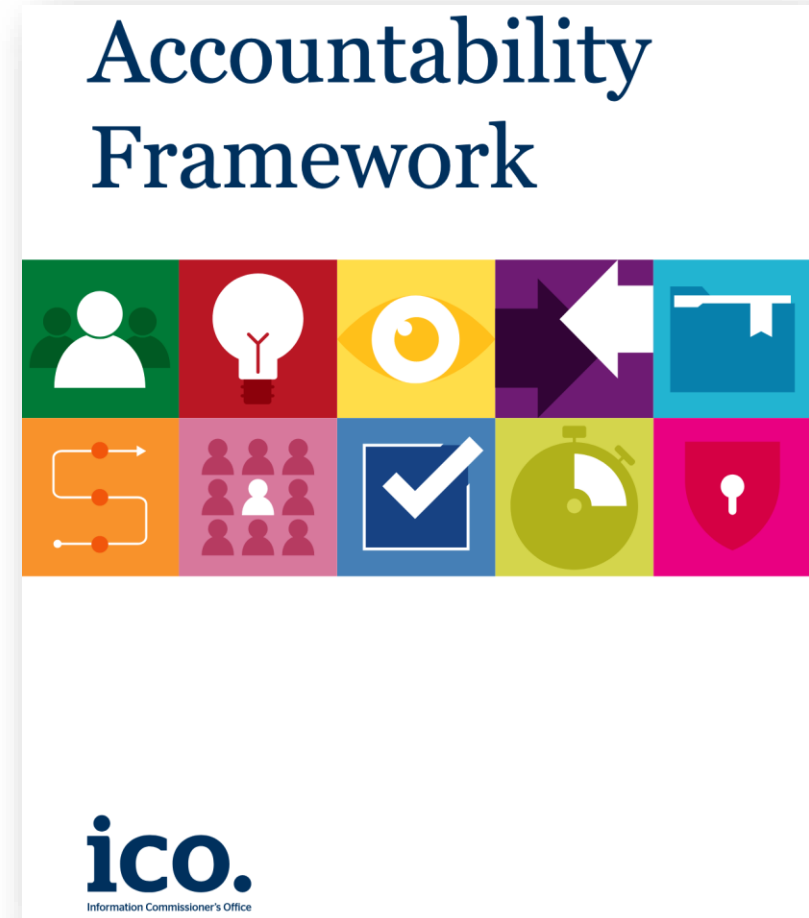
```
1  system:  
2    - fides_key: demo_analytics_system  
3      name: Demo Analytics System  
4      description: A system used for analyzing customer behaviour.  
5      system_type: Service  
6      administrating_department: Engineering  
7      egress:  
8        - fides_key: another_demo_system  
9          type: system  
10         data_categories:  
11           - user.contact  
12      ingress:  
13        - fides_key: yet_another_demo_system  
14          type: system  
15         data_categories:  
16           - user.device.cookie_id  
17      privacy_declarations:  
18        - name: Analyze customer behaviour for improvements.  
19          data_categories:  
20            - user.contact  
21            - user.device.cookie_id  
22          data_use: improve.system  
23          data_subjects:  
24            - customer  
25          egress:  
26            - another_demo_system  
27          ingress:  
28            - yet_another_demo_system
```

All open source! Check them out here: [Privado](#), [Fides Lang](#), [Semgrep](#), [Coconut IDE](#)

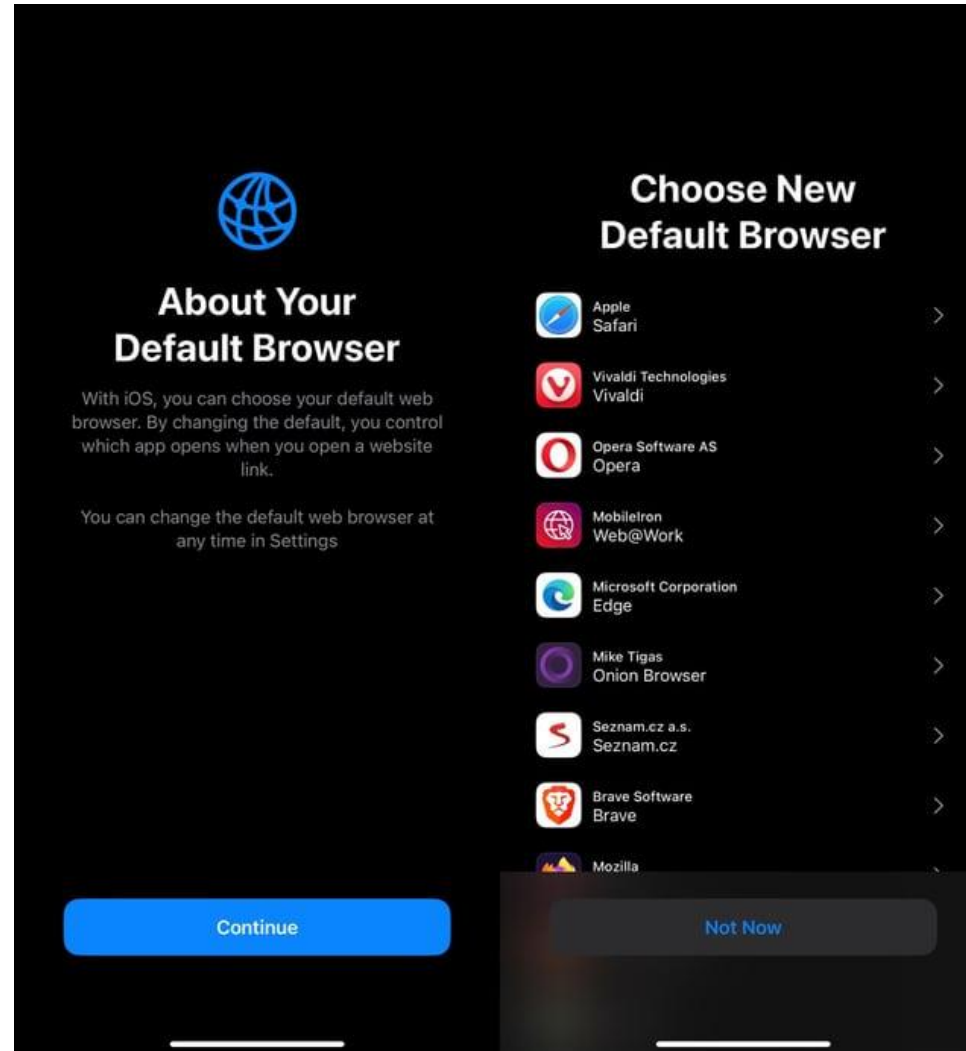


6. Visibility and transparency

- Your users should trust, but verify – give them the info and means to verify you're doing the right thing
- Provide enough info for users to make informed decisions
- Provide complaint and redress mechanisms
- Responsibility for privacy controls should be clearly assigned and communicated



(Un)Informed decisions



(Un)Informed decisions: the reality

We asked people what they think happens after they choose a default browser. **Only about half (52%) of people understand that their default browser is opened when they, for example, click on a link in an email or document.**⁶⁰ This share did not significantly vary across treatment groups, which is surprising as the Q&A screen (displayed to T2-4) explicitly states this is what happens when they select a default browser. Moreover, 6% of people incorrectly believe that they are *only* able to use their default browser (i.e., that all other browsers would be disabled). Even more significantly, **over half (53%) also erroneously believed that their default browser would automatically be pinned to their taskbar.** Previous research suggests this is a common misconception - and the impact of not being pinned to the taskbar, in terms of how this impacts people's usage of browsers, is an area where we consider further research could be undertaken.

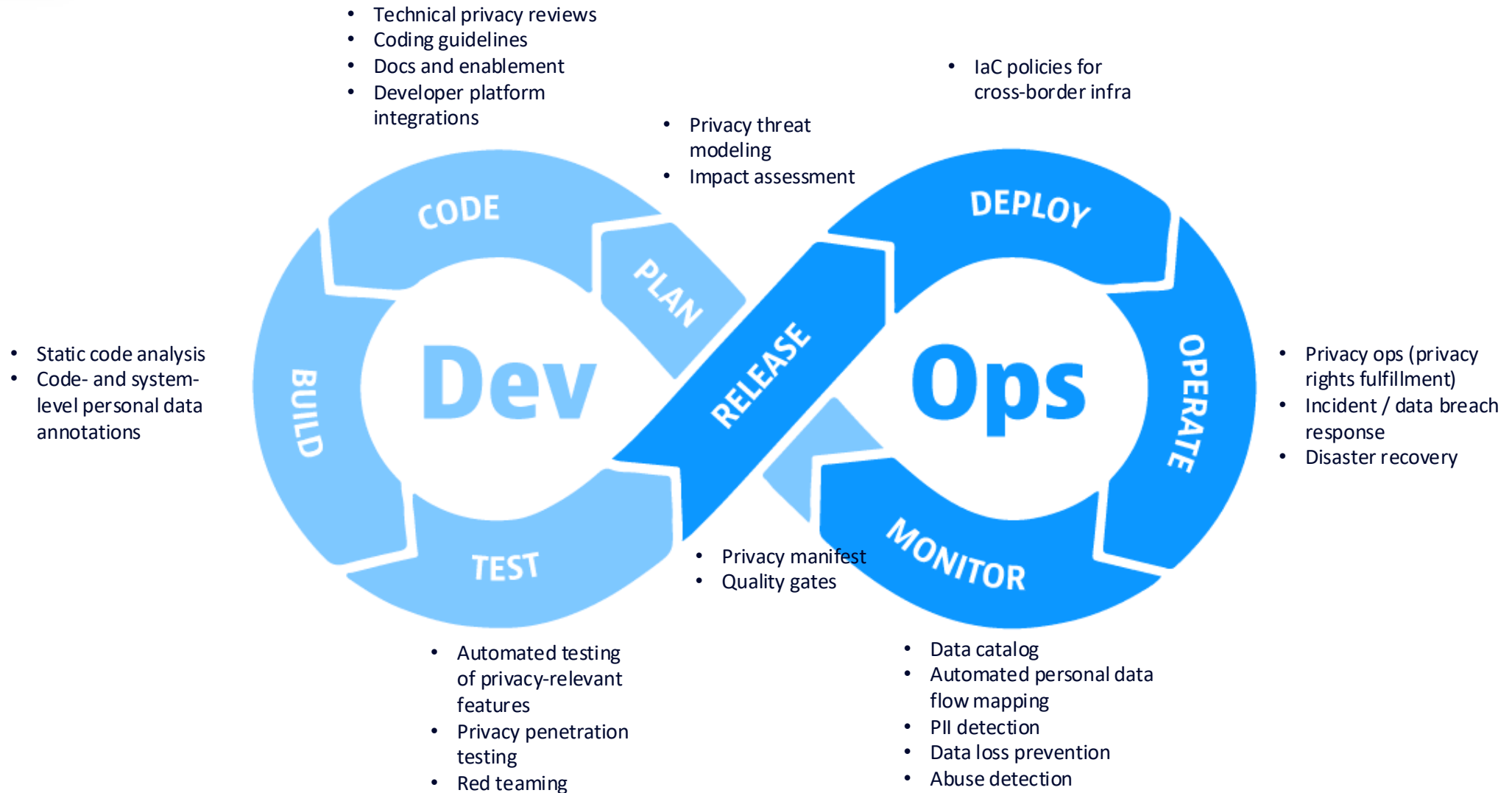
7. Respect for user privacy

- Prioritize the interests of users
 - Including those who might be indirectly affected (e.g. B2B products)
 - Get informed consent to process data
- Design-in support for privacy rights
- Build human-centric systems - with human support available!

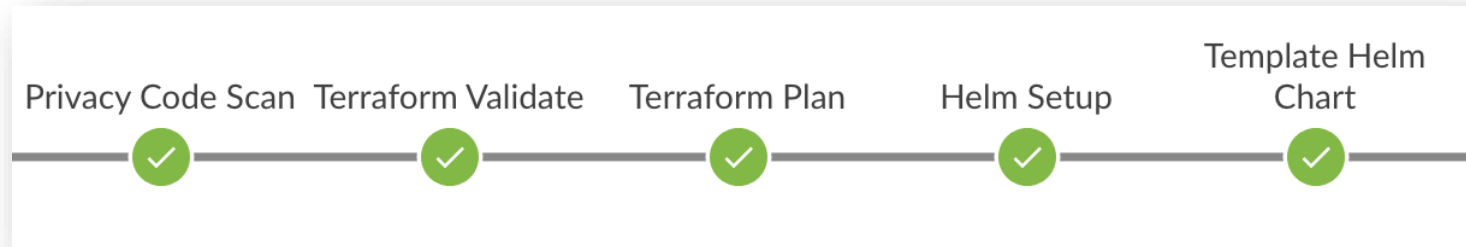


When?

Privacy by design in the SDLC with privacy controls



Build stage: privacy code scanning



git-privacy-ops 31 May 2024


Personal data should not be output to logs

Please confirm that:

- Processing this personal data is **relevant** and **necessary**
- If this data is persisted, a retention period is enforced
- You have followed the [Privacy Coding Guidelines](#)


Reply 🗨️ ...


Release stage: privacy manifest quality gate





Data Linked to You


The following data may be collected and linked to your identity:


 Purchases


 Location


 Contact Info


 Contacts

 User Content


 Identifiers


 Usage Data


 Diagnostics


 Other Data


Analytics


 Purchases
Purchase History


 Location
Coarse Location


 Contact Info
Email Address

 User Content
Gameplay Content
Customer Support

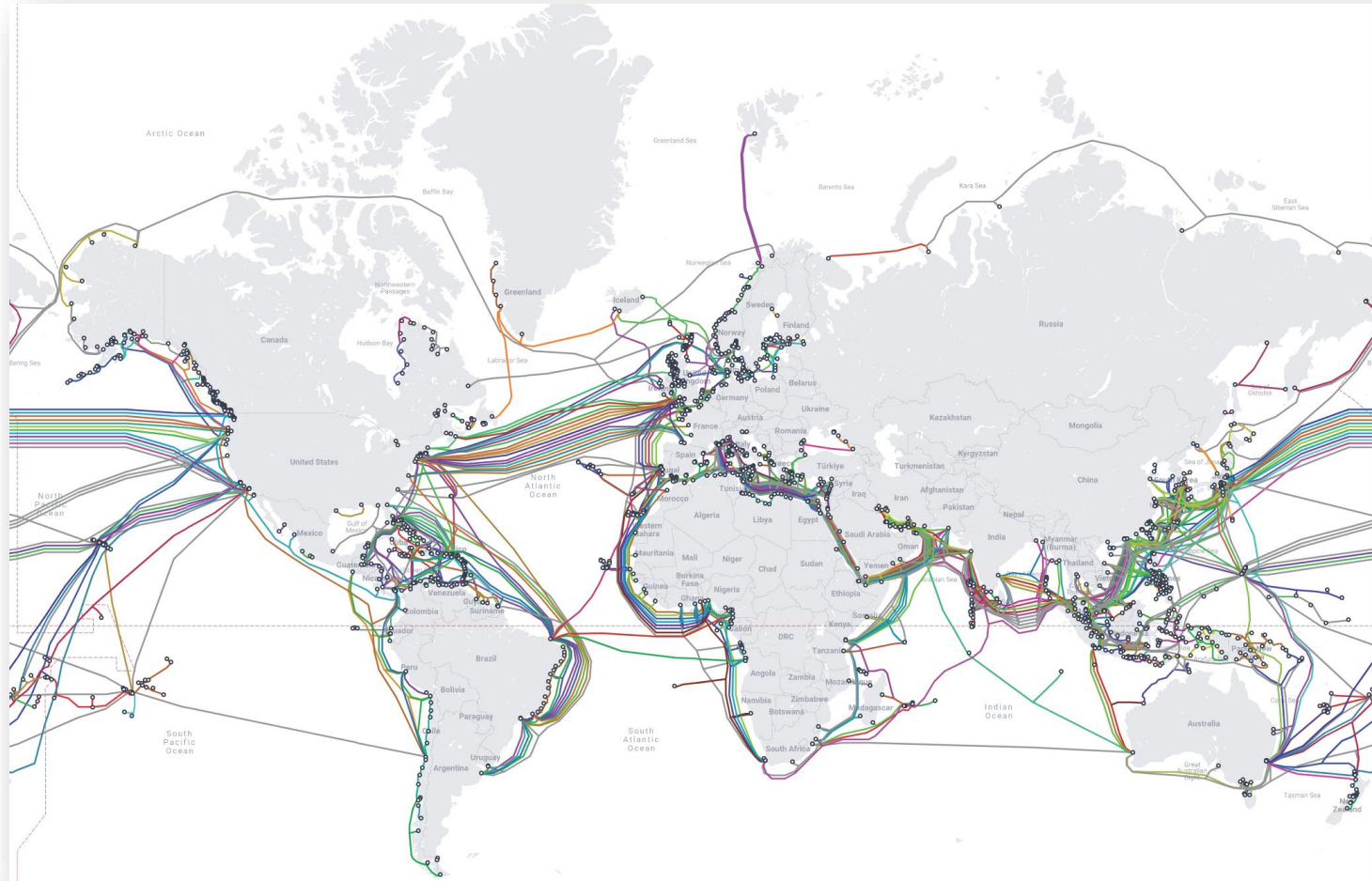
 Identifiers
User ID
Device ID

 Usage Data
Product Interaction
Advertising Data
Other Usage Data

 Diagnostics
Crash Data
Performance Data
Other Diagnostic Data

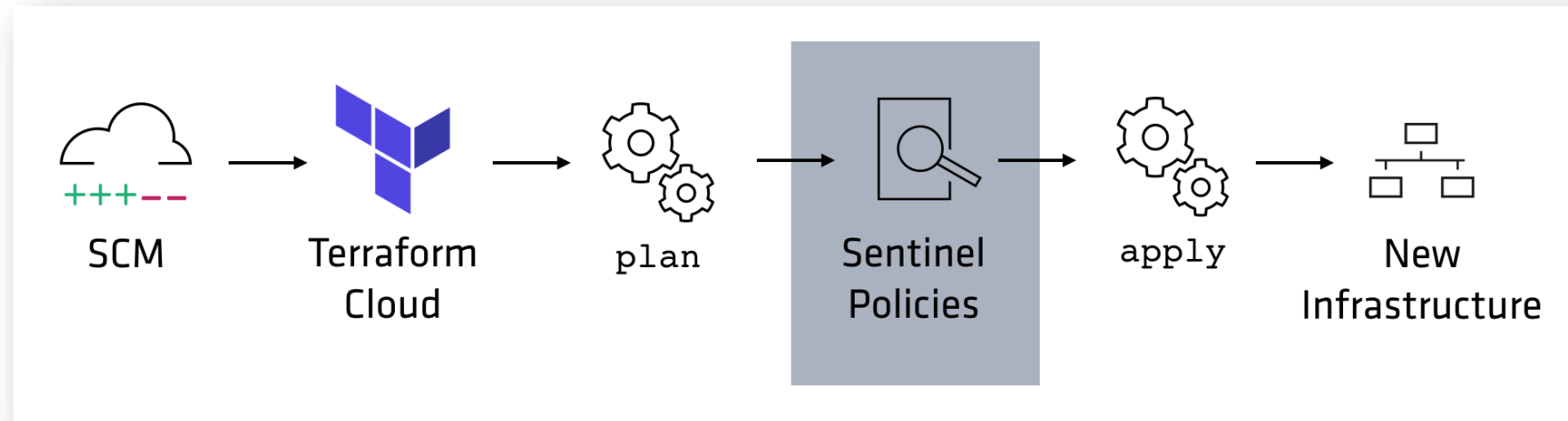
 Other Data
Other Data Types

Cross-border data transfers in the cloud



Source: [Submarine Cable Map by TeleGeography](#)

Deploy stage: IaC policies for cross-border infra



Deploy stage: IaC policies for cross-border infra

```
1  import "tfplan-functions" as plan
2
3  allowed_zones = [
4      "eu-central-1a", "eu-central-1b", "eu-central-1c"
5  ]
6
7  # Get all EC2 instances
8  allEC2Instances = plan.find_resources("aws_instance")
9
10 # Filter to EC2 instances with violations
11 violatingEC2Instances = plan.filter_attribute_not_in_list(allEC2Instances,
12     "availability_zone", allowed_zones, true)
13
14 main = rule {
15     length(violatingEC2Instances["messages"]) is 0
16 }
```

Deploy stage: IaC policies for cross-border infra

If not using HCP Terraform:

[Open Policy Agent](#)



Thanks for listening!

