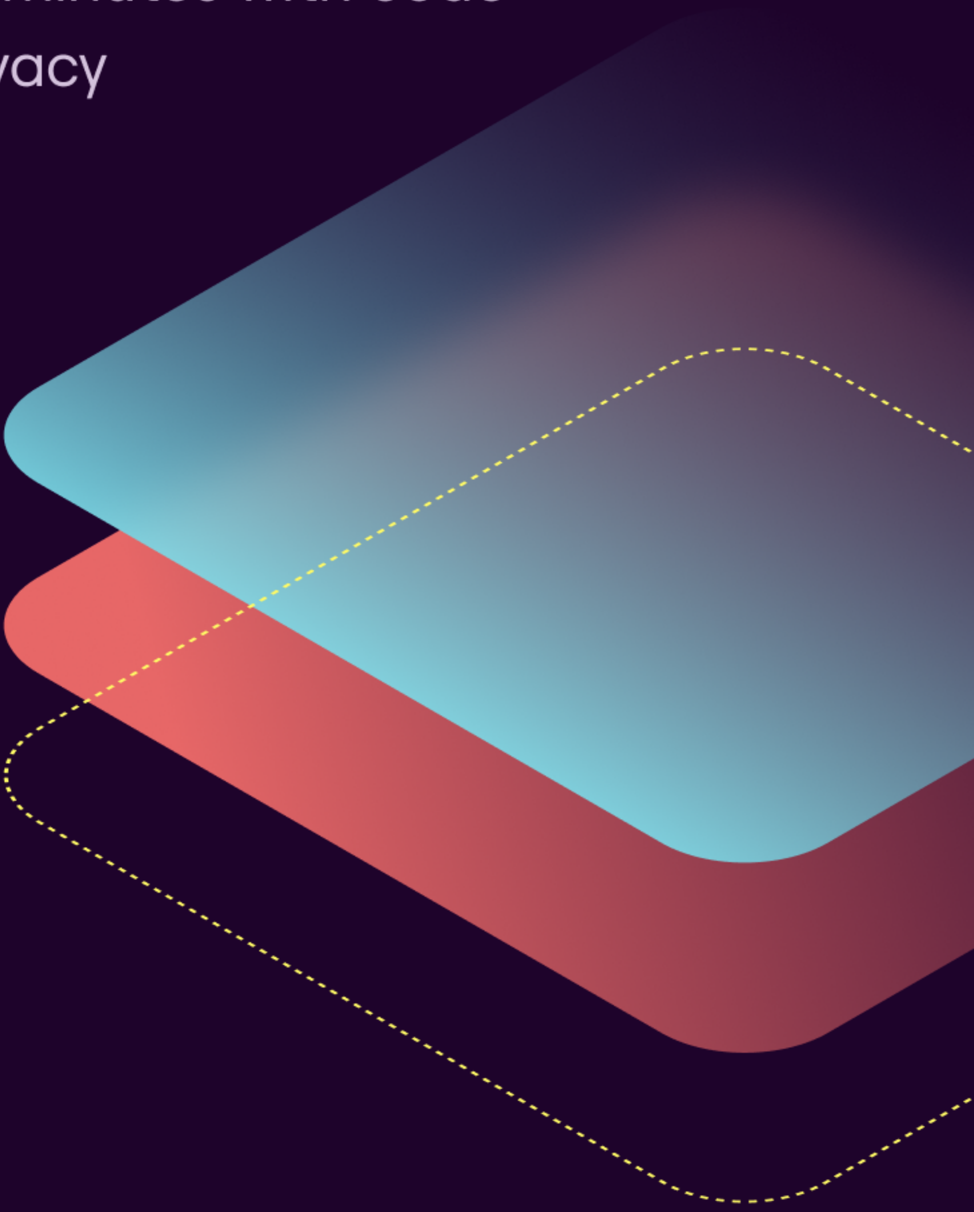
privado

Data Mapping
A New Approach

From months to minutes with code scanning for privacy



AUTHOR

Vaibhav Antil

CEO, Privado.ai

For more resources. Visit Privado.ai

Index

Section: 1
Overview & Objective

Overview & Objective

Data mapping is the backbone of a privacy and security program, but when it comes to mapping engineering processes, these maps are often outdated and incomplete. This is because privacy is considered a legal issue rather than a technology challenge.

This publication discusses the challenges in operationalizing data mapping for engineering, including the gap between privacy and engineering teams and how adopting the ‘application-led’ approach to data mapping results in more accurate and up-to-date data maps.

Executive Summary

Data mapping is a topic of discussion in the security and privacy community because, even after all these years, the process of creating data maps is still manual.

- Interviewing teams to map business processes
- Running assessments to figure out the data journey
- Reviewing legal basis and adding privacy controls
- Creating privacy reports using the analysis

- But by the time the data is collected and processed, it's outdated by several months, making the reports an unreliable source on which to base business decisions.

- By the time the reports are prepared, the data collected in the exercise is outdated by several months, and there may be new processes and 3rd parties which might be processing data. Using outdated resorts as a starting point may result in an inaccurate representation of privacy controls and processes in privacy policy statements. Inaccurate representation may result in non-compliance with the law, a fine, or a lawsuit by the regulator.

- To bridge the gap with the engineering team and produce accurate data maps, initiate the process by analyzing the application code to find the sources where data is collected and later approach engineering colleagues to clarify.

- By analyzing application code, the data map will have complete coverage of the repositories and tools built by engineers directly from source code tools. This approach to data mapping helps security and privacy teams to align with software sprints. So as new code is taken live, it can be quickly analyzed for any privacy violation and prevented from going into production if it is not adhering to the promises made in the privacy policies.

Section: 2
Why talk about data mapping today?

Why talk about data mapping today?

Privacy and security professionals started discussing means to map the flow of data in their organization in early 2016 when laws like GDPR were in the early stages of adoption.

> *The reason data mapping is still relevant in 2022 is that we're dealing with it in the same way we were dealing with it in 2016.*

Stephen Davis,
CISO

THRASIO

Even after all these years, the process of creating data maps is still manual. The privacy and security team starts by sending out questionnaires to their colleagues across the organization to collect data on the various team's business processes. They follow this by setting up interviews, conducting a privacy review after collecting all the data, and finish the process by creating a report on their findings.

With this process, by the time the team gets the data they are looking for, it is very likely that it will be outdated. Thus, making the reports an unreliable source for basing business decisions on.

There are some tools available in the market that suggest that they automate data mapping, but these tools generally can only automate a small portion of the workflow because they are based on business processes. So, while there can be some automation at that level, it won't be sufficient at the scale and speed that modern organizations with agile product development cycles and thousands of code repositories will need. This leaves the heavy lifting to the teams trying to get visibility of data flows.

> *I've yet to find one [a tool] that actually automates the creation of data maps in a meaningful way without really needing months of manual effort.*

Stephen Davis,
CISO

THRASIO

Section: 3

Current approach to data mapping

Current approach to data mapping

> ***On average it takes most organizations 6-12 months of time to do data mapping***

Most privacy and security professionals follow a business-process-led approach to data mapping. In it, data maps are created in three steps:

Step 1: Interviews
The data mapping process starts with interviewing business heads from across functions in the organization to understand the key processes involved in the team's day-to-day tasks.

After getting a big picture, the team converses with other stakeholders in the function to get better insights into the processes. The step generally takes a couple of weeks and involves a lot of back-forth between colleagues from across functions and manual data processing.

Step 2: Assessment

After collecting questionnaires and interviewing colleagues, the next step in the data mapping process is running assessments to determine the data's journey in their organization.

In this step, questions such as: What data is collected? Where is it going for storage? Where is it being shared? Who's accessing the data? are answered.

This step generally takes under a week to complete and needs to be thoroughly vetted by the privacy/security team before moving on to the next step.

Step 3: Privacy Review

After building the data inventory, it needs to be analyzed to figure out the legal basis for data collection for the organization, to chart out the privacy controls needed, and add them to the business processes.

If everything runs smoothly and the team is able to create data processing policies and add controls to business processes, now is the time to create privacy reports.

> *The three-step approach to data mapping is the ideal state we follow today, but in my experience, what is ideal, and what tends to be -- are two very different things in the world of privacy and security.* ***Getting to this ideal stage can take upwards of six or even 12-plus months.***
>
> *It's frustrating for us privacy professionals because we spend half our time chasing down documentation. And even more frustrating is that when we get the documentation, it's often outdated. So we know that we're working without accurate data, and depending on how much change may or may not exist within the organization, the outdated information could be woefully outdated.*
>
> *To try and make decisions off of that is nearly impossible*

Stephen Davis,
CISO

THRASIO

Section: 4
Gap between privacy and engineering teams

The gap between engineering and security/ privacy teams

There is a fundamental difference in the way engineering teams function and release their work to how legal and business processes work. This exacerbates the challenge of charting data flows for privacy and security teams.

Applications built today are complex in structure with distributed architectures, 3rd party integrations, and multiple microservices connected via APIs. New product changes are pushed daily and monitoring whether any of these changes affects sensitive data becomes unfeasible under a business-process approach to data mapping. This results in inaccurate data maps, which can lead to misrepresentation of privacy controls and make the organization susceptible to fines, and lawsuits.

> *It's not like security/privacy professionals or engineers are actively trying to misinterpret the data getting captured; it's just that they don't know what they don't know. And because developers and business teams speak different languages, a lot gets left in translation.*

Stephen Davis,
CISO

THRASIO

Here are the key challenges security/privacy teams face working with engineering teams:

Challenge 1: Lost in translation
Engineers talk in the language of applications, systems, services, and APIs. So whenever the security/privacy teams try to converse with them in technical legal language — it's almost like two people are talking in different languages, and a lot of context is lost in that translation.

Challenge 2: Validating controls
We generally work with the engineering teams to build privacy controls in their processes before product development starts. But often, what we plan doesn't end up being in the final product -- and we barely get time to validate the controls present in the code before it's pushed into production.

Challenge 3: Engineering processes are decentralized

Engineering teams are decentralized. Every team decides its tech stack, the kind of database the product will use, and which third parties it will work with, and since each team makes different decisions, which have its own privacy risks, and challenges.

Challenge 4: Code is updated often

Say the team does all the grunt work and maps every process inside the product. The engineering team will push new code every other week, so the team's understanding will get outdated quickly—and the data maps will be of no use. So the top-down privacy commitments made in the privacy policy and what's in the code will not match.

Section: 5
Code scanning approach to data mapping

Application-led approach to data mapping

Engineers write code day in, day out. **All of the information about how data is collected, shared and used is readily available in the code repositories.**

Organizations can bridge the gap between security/privacy teams and engineering teams by initiating the data mapping process with an analysis of the application code and later approaching engineering colleagues to clarify information.

> *When the team starts with code, it can go directly to the engineering colleagues and ask, "Hey, this is the application that you're working on; I see you have sign-up forms inside this repository; what personal data do you process or collect?" They will know what is being talked about because they're being asked precise questions about the code they've written.*

Vaibhav Antil,
CEO

privado

Analyzing application code can yield complete coverage of the repositories and tools built by engineers directly from source code that lives in GitHub, GitLab, and Bitbucket, where code is maintained. This data can be used to build an application catalog to help security and privacy teams collaborate with others.

And finally, the team can align privacy reviews with software sprints. So as new code is pushed to production, it can quickly be analyzed for any privacy violation and prevented from going into production if it is not aligned with the promises made in the privacy policies.

> *If security/privacy teams shift approach from a business process approach to an application approach, they get the advantage of being close to the source of truth right out of the door.*

Vaibhav Antil,
CEO

privado

That said, an approach focused on analyzing application code will still present challenges:

> *Building data maps with an application-led approach is not a silver bullet, because even if it's used, it will still present two challenges, [1] automation and [2] scale.*
>
> *And the reason is that while you're speaking the same language, making data maps will still be manual. You still have to ask your engineering colleagues, "Hey, what is your application doing?" which means you will spend countless hours both on the development and the privacy team sides; the final map will still rely on human judgment. So, you will still need to double-check if the representation is accurate.*
>
> *And again, making data maps with this approach will not scale because of the agile way of product development; code is pushed to production every other week. So these application data maps will need to be updated, as well, really, really fast.*

Vaibhav Antil,
CEO

privado

And that's where Privado's privacy code scanning tool adds value.

How Privado helps teams build dynamic data maps in real-time

Application-led approach via code-scanning and automation for Privacy

How Privado helps teams build dynamic data maps in real-time

An application-led approach is insufficient when it lacks automation and git integration. Privado is a static privacy code scanning platform that connects to the application's source code through management tools like Github, Gitlab, and Bitbucket. Privado then runs a static code scan on the entire source code [or select repositories] to catalog all the applications inside the code to build dynamic data maps by pulling real-time information from the code in less than 30 minutes.

Since Privado integrates directly with the sources where the application code is stored—where data collection, sharing, and processing decisions are made—its privacy code can automatically build dynamic data maps without requiring any extra effort.

The data map charts everything from data elements, third parties, databases, APIs, and data flows; and automatically updates as and when the code changes—ensuring there's always an up-to-date view of the data flows inside the product available for use.

And because the data maps are accurate and dynamic, the security/privacy teams can confidently build privacy policies with accurate representations with it.

Later, if the code goes outside the boundary of this privacy policy, the team gets an instant alert that lets it either update the policy or stop the change. Plus, the tool builds an inventory for privacy operations that helps the team keep track of all the databases and third parties where application data flows—which the team can analyze to implement privacy control using policies and apply data minimization limitations.

About Privado

At Privado, we have a core belief that people deeply care about safeguarding their privacy from governments and corporations alike and that trust and privacy are critical for the internet economy to function. Our mission is to empower developers to create privacy-first products that prevent harm to people, comply with global data protection laws, and enable trust from stakeholders.

Our static code analysis platform lets you track whether your developers are sharing personal data across third-party apps and what are those data types. Schedule a demo to learn how we track personal data flows to third-parties.

Great brands trust us:

headspace

here

ZEGO

THRASIO

trendyol

Terra.do

asos
discover fashion online

BlackBerry

ibis
HOTELS

meesho

Contact:

✉ hello@privado.ai

@PrivadoHQ

in Privado.ai