rudderstack

Whitepaper

Implementing Privacy for Compliance

Guide: Implement privacy control in customer data pipelines

Customer data pipelines play a critical role in the privacy of your customer data. They are one of the primary and most expansive collectors of your customers' personally identifiable information (PII). They are also one of the most expansive sharers of customer data - with one of the primary use cases being event streaming to frequently large libraries of destination integrations.

Due to their specialized role of collecting and sharing customer data, customer data pipelines can either help ensure your data privacy or wreak havoc on it.

This guide will explain how your customer data pipeline can help improve your data privacy and how to ensure your data privacy with RudderStack.

DATA PRIVACY VS DATA SECURITY

To remove one common vector of confusion before we launch into this post, we want to make sure the difference between data privacy and data security is clear.

Data privacy - the focus of this post - is about what data is collected, stored, how long it is retained, and what customer data is shared. *What customer data are you collecting and storing, and how are you using that data?*

Data security is about how collected data is protected - where data is stored, who has access, whether data is encrypted, etc. *How are you keeping the customer data you store safe?*

YOUR CUSTOMER DATA PIPELINE CAN IMPROVE YOUR DATA PRIVACY

Your customer data pipeline can give you fine-grained control over what data you are sending to which tools, and what data you are storing. This can help you avoid data privacy issues before they ever occur.

The three processes below are designed to help you ensure that your customer data stays private. You can implement all of them with a robust customer data pipeline tool like RudderStack.

Data Masking

Data masking is taking fields in your event data and obfuscating them. This is most frequently used to hide PII. Your customer data pipeline can mask your PII before it is ever sent to a destination or stored in your warehouse.

For example, if your event payload includes the following attributes...

```
"globalUserId": "XYJ458907432AAC",
"userId": "contactUser",
"userFirstName": "Rudder",
"userLastName": "Stack",
"userEmail": "contact@rudderstack.com",
"userSSN": "123-45-6789",
"eventType": "newsletter-sign-up"
```

One level of data masking would remove the directly identifiable PII, like SSN and email address.

```
"globalUserId": "XYJ458907432AAC",
"userId": "contactUser",
"userFirstName": "Rudder",
"userLastName": "Stack",
"userEmail": "XXXXXXX@XXXXXXXXXXX.XXX",
"userSSN": "XXX-XX-XXXX",
"eventType": "newsletter-sign-up"
```

Another more stringent level of data masking would remove all unnecessary attributes. Since most of the attributes in this payload are identifiers in one way or another, only the global identifier and event type would be unmasked.

```
"globalUserId": "XYJ458907432AAC",
"userId": "XXXXXXXXXXX",
"userFirstName": "XXXXXX",
"userLastName": "XXXXX",
"userEmail": "XXXXXXX@XXXXXXXXXXX.XXX",
"userSSN": "XXX-XX-XXXX",
"eventType": "newsletter-sign-up"
```

Attribute Removal

Similar to data masking, attribute removal is selectively removing attributes from your event data. Not every application you send event data to needs all of the customer data you collect in your events. Attribute removal can be used to remove PII or to remove unnecessary customer data and reduce payload.

Using the same event example, if you wanted to activate that data by triggering an email send in your email/marketing automation tool, you would remove the unnecessary attributes for sending an email - userId and userSSN.

```
"globalUserId": "XYJ458907432AAC",
"userFirstName": "Rudder",
"userLastName": "Stack",
"userEmail": "contact@rudderstack.com",
"eventType": "newsletter-sign-up"
```

Event Filtering

Not all of the tools you stream events to need every type of event. Event filtering is the process of removing events from an event stream based on filtering criteria. This ensures that only the events you want to activate on are ever shared with the tools you activate. So you don't overshare your customer data with tools that only use a small portion of it.

Using the same event example, if you filtered to where eventType = "newsletter-sign-up," the sample event would be included. If you filtered to where eventType != "newsletter-sign-up," the sample event would be excluded.

RUDDERSTACK TRANSFORMATIONS KEEPS YOUR CUSTOMER DATA PRIVATE

RudderStack Transformations allows you to transform your event data in-flight - after collection, before delivery. Transformations are reusable functions - written in JavaScript - that can be applied to the data in your event streams before delivery to a destination tool or your data warehouse.

With RudderStack Transformations, you can implement all three of the data privacy processes detailed above, plus any other type of data transformation you can code in JavaScript. Transformations are applied on a destination-by-destination basis, so you can implement specific privacy processes for each tool you use and your data warehouse - only sharing the exact

customer data you need to share. And they are reusable, so it's easy to apply the same transformation to multiple destinations. Write it once and apply it everywhere.

We maintain an open-source repository of Transformations templates that implement a wide variety of data transformations - from data masking, attribute removal, and event filtering to event enrichment. The JavaScript code for individual transformations is stored in this repo. You can copy it, edit it to work with your data, and paste it into RudderStack Transformations.

- Data masking template
- Attribute removal template
- Event filtering template

If you want more details about using RudderStack Transformations, read our step-by-step guide on adding custom Transformations.

If you want more details about how to mask PII with RudderStack Transformations, read our blog post Protect Personally Identifiable Information (PII) in Your Apps Using RudderStack.

SIGN UP FOR FREE AND START SENDING DATA

Test out our event stream, ELT, and reverse-ETL pipelines. Use our HTTP source to send data in less than 5 minutes, or install one of our 12 SDKs in your website or app. Get started.

RudderStack is the warehouse-first, customer data platform built for developers.

We take a new approach to building and operating your customer data infrastructure, making it easy to collect, unify, transform, and store customer data as well as securely route it to a wide range of marketing, analytics, sales, and product tools.

Over 18,000 sites and apps run RudderStack including Crate & Barrel, Acorns, Hinge, Stripe, Allbirds, and more.

rudderstack.com

@rudderstack