twilio segment

Why bad data happens to good companies

Drive revenue

Elevate customer experiences

Introduction

Data is many things: a competitive differentiator, an asset, a barometer for decision-making. Data is what's training AI models and unlocking advanced analytics. But when it isn't properly managed, all those positive attributes can flip – turning data into a false lead, legal liability, or financial burden.

Investing in good data seems like a given, but it's much harder to achieve than you may think. Just collecting data from the various different apps and tools in your tech stack – and making sure this data is complete, accurate, and accessible across teams – is becoming harder every year. Then there's the matter of actually being able to use this data in day-to-day operations.

Bad data comes in different forms: it can be stale, inaccessible, or untrustworthy. But one constant remains: bad data leads to bad outcomes, every time. Whether it's costing businesses millions of dollars each year, leading to lackluster customer experiences, or eroding brand reputation.

Let's discuss how to achieve good data within your organization, and how to avoid the pitfalls of bad data.

Table of Contents

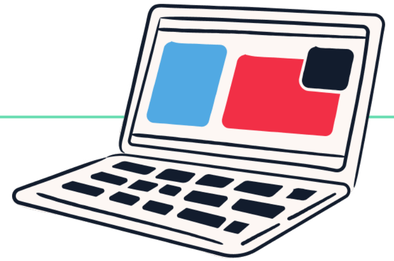
What is good data?	**03**
What can you achieve with good data?	04
What causes bad data?	**05**
Data Silos	06
Lack of standardization	08
Incomplete customer view	10
Stale data	11
Not complying with privacy regulations	13
Protect data quality at scale with Twilio Segment	**15**

Chapter 01

What is **good data?**

First, let's define what we mean by "good" data.

Good data is accurate, consistent across your tech stack, compliant with privacy regulations, and available to the teams who need it. In short, good data is high in quality and effectively managed across its lifecycle – from how it's collected, to how it's stored, used, and eventually deleted.



What is good data?

- Gathered from trustworthy sources (e.g., don't buy customer data from shady third-party vendors - or, ideally, at all)
- Enriched (e.g., integrating real-time and historical data for deeper insights)
- Unique (e.g., no double entries for the same event)
- Timely (e.g., available when it's needed and updated in real time or near real time).
- Complete (e.g., no missing values)
- Consistent (e.g., reporting doesn't vary between different tools)

What can you achieve with good data?

Good data unlocks key insights for your business: like how to streamline operations, what should be prioritized in your product roadmap, and which acquisition strategies have the highest ROI.

The common thread here is good data's ability to generate more revenue, whether by mirroring customer expectations, finding the right product-market fit, or not wasting time and resources on channels that generate little returns.

Here are a few examples of what can be achieved with good data:

- Leading online classifieds provider **Adevinta** created over 1,290 audience segments using consolidated, real-time data, and saved €190K a year on marketing campaign costs due to better targeting and personalization.
- Retailer **Veronica Beard** reduced customer acquisition costs by 20% and increased ad spend by 11% after creating a single source of truth for their data and excluding "low-value customers" from ad campaigns.
- Global healthcare company **Sanofi** saw a 95% increase in efficiency after moving to a real-time, event-driven architecture that gave teams access to customer data instantly, and allowed them to enrich customer profiles in their warehouse. (Before, it would take them roughly three days to activate data.)

Chapter 02

What causes bad data? (And how to fix it)

There are a few common culprits when it comes to bad data, from siloed systems to a lack of consistent naming conventions. As we cover these in more detail, you'll see an underlying point emerge: preventing bad data starts from the ground up – from the data governance policy you implement to the interoperability of your tech stack.

Data silos

Data silos are easy to create and incredibly annoying to fix. The average business has hundreds of tools in their tech stack (**for a typical mid-market company there's roughly 185 apps used throughout their workflows**). And while each team has access to the data they collect in their owned tools, it can be difficult to share it across an organization unless it's manually pulled and formatted by an engineer.

In one study of 400+ decision-makers across industries, **75% said they struggled with siloed data**. When discussing why they felt their current data and analytics management wasn't helping them evolve with customer expectations, three main reasons emerged:

- A lack of integrations
- Difficulty in setting up integrations
- An inability to handle the volume of data they had to collect and unify

For businesses to stay adaptable, they need to think about the interoperability between their tools and systems (especially as their tech stacks evolve).

An overview of some common sources of customer data

Website
Javascript or other website tagging option

Mobile
iOS, Android, AMP, etc.

Servers
NET, Clojure, Go, Java, Node.js, PHP, Pixels, Pythons, Ruby, etc.

CRM systems
Salesforce, HubSpot, etc.

Payments systems
Stripe, Amazon Payments, etc.

Attribution Platforms
AppsFlyer, Adjust, Kochava, Tune, etc.

Email systems
MailChimp, SendGrid, Marketo, etc.

Advertising campaigns
Facebook, Google AdWords, etc.

Help desk systems
Zendesk, Twilio Flex, Salesforce Service Cloud, etc.

75%
struggled with siloed data

How to fix it

Data integration is the process of combining data that's otherwise scattered across different tools and systems. One of the overarching points of this is to create a central repository for your data (usually in a data warehouse or a data lake), which functions as a source of truth for businesses.

Choosing the right data integration strategy for your business will depend on a couple of factors, like:

- Scalability (e.g., how will you handle changes in your tech stack?)
- Cost and time commitment (e.g., manually building and maintaining ETL pipelines will require more resources from your engineering team)
- Security (i.e., how are you protecting data from both a legal and ethical standpoint)
- Data governance (i.e., how are you guaranteeing the integrity of your data at scale)

Twilio Segment offers hundreds of **pre-built integrations** to help with this, allowing businesses to connect tools and apps in a matter of minutes (or **build custom Sources and Destinations** as needed). You can also use **replay** to send a limited sample of your data to a new tool to test it out before committing, which can help avoid vendor lock-in.

Lack of standardization

Data standardization is the process of transforming data into a uniform format or structure, to ensure consistency and compatibility between different systems and datasets.

When an organization is working in silos, data discrepancies can run rampant. Let's use the example of a SaaS company whose app runs on web, iOS, and Android. Without data standards, different teams start to measure the same event in different ways. Even something like slightly different spellings, hyphenation, property names, or values can wreak havoc on data collection and analysis – causing the same event to be counted multiple times.

In the chart below, you might notice that:

- Website and Android use spaces in event names, while iOS uses hyphens.
- Website and iOS use camelCased property names, while Android uses snake_case.
- Website uses lowercase property values, while iOS uses Title Case and Android uses Title Case or integers.

As a result of these inconsistencies, there's no way to accurately compare the same event across platforms.

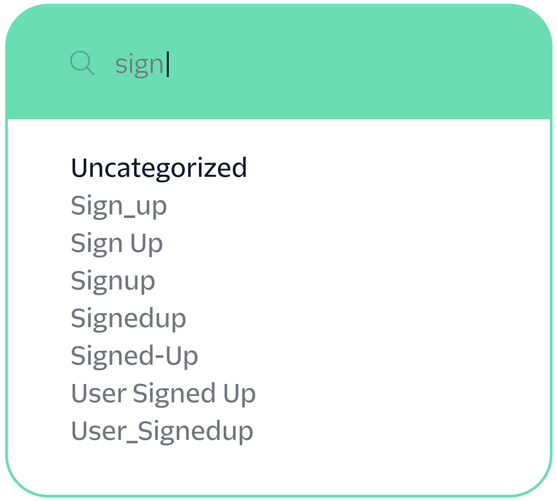
Website	iOS	Android
Signed In referralType: organic	Signed-In referralType: Organic	Signed In referral_type: Organic
Step Completed stepName: one	Step-Completed stepName: One	Step Completed step_name: 1

How to fix it

To fix data discrepancies, it's important to first align everyone around a **single tracking plan** and standardized naming conventions.

Your tracking plan will outline what events you'll be tracking, where in your codebase/app they will be tracked, and why you're tracking these events (e.g., what business goal is this tied to?) Here's a customizable **tracking plan** to help get you started.

Your naming conventions will ensure that there's a uniform format for how you refer to the data being tracked (e.g., "user_signup" vs. "User SignedUp").



For naming conventions, we recommend using an object-action framework. The idea is simple: first, choose your objects, or the key "pieces" of your app and website that customers interact with. For an e-commerce company, these objects might be "product," "cart," or "promotion." And for a SaaS company, they might be, "account" or "workspace."

Then, choose your actions. This is how people interact with those objects. So, with "product," a common action might be "viewed," which would then become, "product_viewed."

You also want to be crystal clear about the casing. This might seem nitpicky, but it's imperative in the long run. Here are the five most common options:

- all lowercase – account created
- snake_case – account_created
- Proper Case – Account Created
- camelCase – accountCreated
- Sentence case – Account created

Incomplete customer view

Issues like data silos and data inconsistencies can create another roadblock for businesses: an incomplete customer view.

Customers today are constantly switching between devices and channels as they interact with a brand, making it increasingly difficult for companies to attribute the right interaction to the right person.

It's impossible to know your customer if you don't know what they're doing. To safeguard against this, businesses need to merge the complete history of a customer's interactions – across web, mobile, email, etc. – into a unified profile. That is, they need to perform identity resolution. This is what will empower businesses to deliver personalized, one-to-one experiences, spot churn risks, and calculate average lifetime value.

However, doing identity resolution entirely in-house is a huge undertaking. While it is possible to create a basic version of an ID graph using SQL and the data in your warehouse, even if you're a SQL genius you'll constantly be re-tweaking data models. (Or finding new ways to join data tables from completely disparate solutions.)

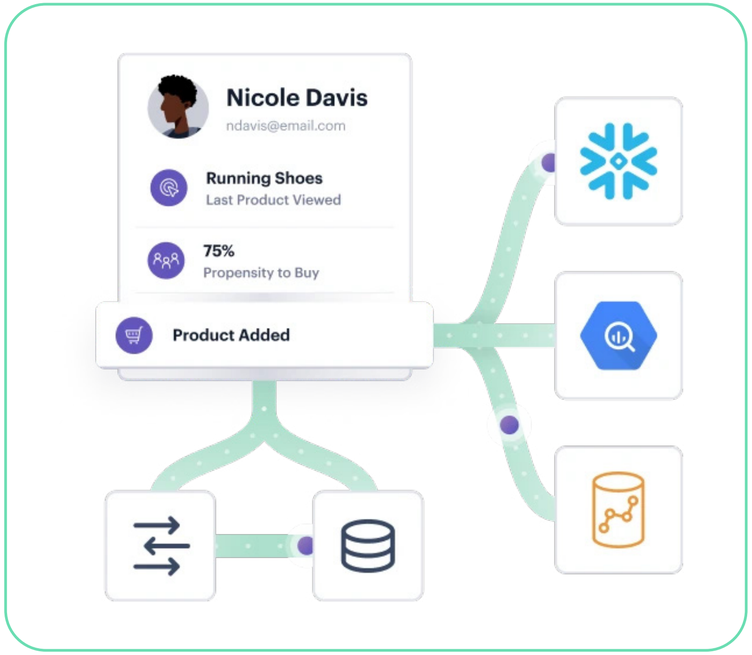
Then there's the matter of making those ID-resolved profiles available in downstream marketing, sales, and customer success tools for activation. This will require more complicated data pipelines and custom code for connecting new systems. You'll also have to maintain the ID graph over time, which means managing audit cycles, navigating legal and regulatory compliance, and ensuring data security.

How to fix it

With an API-first platform like Segment, businesses can offload the heavy lifting involved in identity resolution.

Segment's **identity graph** can stitch together user behavior from every channel (including online and offline touchpoints) in real time. It also can attribute anonymous behavior to a known profile once a user is identified in your database. And for B2B companies, it can generate a graph of the relationships between users and accounts.

Even better, the ID graph is customizable. You can bring in your own external IDs and customize which identifiers and sources cause associations. Segment can also **send these customer profiles to your warehouse so they're enriched with historical data**, and then send them back to downstream tools so they can fuel campaigns.



Stale data

Stale data refers to information that is outdated and no longer accurate or relevant. It occurs for a few reasons: a lack of integration between systems, delays in data processing, and natural decay (e.g., a customer moves and their shipping address is no longer correct).

Regular monitoring and maintenance of datasets is essential to ensure data remains accurate and relevant. However, businesses also need to think about how they're updating datasets in real time to keep pace with evolving customer journeys.

Take a customer who recently completed a purchase: **they should be removed from your ad targeting lists immediately** to avoid wasting ad spend. Then there's fraud detection, which depends on near instantaneous detection and prevention.

However, businesses need to have the right architecture in place to handle the high velocity and volume of this streaming data, and the ability to validate data as it flows through these systems to protect its integrity.

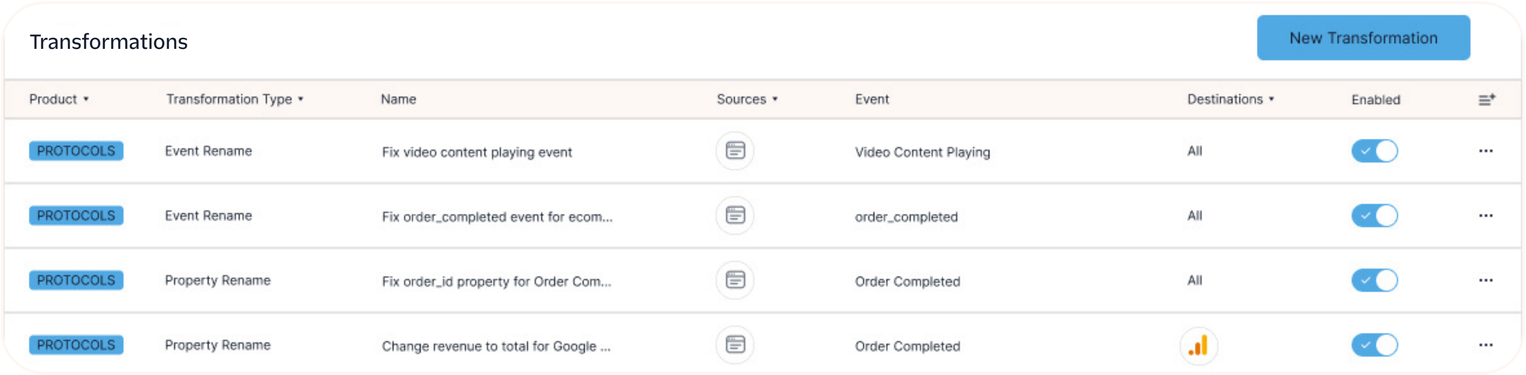
How to fix it

One way to prevent stale data is to embrace real-time data processing. We've seen this with the rise of **event-driven architectures**, which uses "events" to communicate between applications. (Events can be user-generated, like a customer placing an e-commerce order, or system-generated, like a triggered subscription payment reminder.)

One thing to note here is that you need to be validating data at the same rate it's being processed. With Segment, we're able to flag any data that doesn't match your predefined tracking plan so it can quickly be reviewed and remedied. In fact, with Transformations, teams can actually **transform "bad" data as it flows through the pipeline** (e.g., changing event or property names, or customizing data for a specific destination).

Another important caveat is that not all data needs to be processed in real time. Companies need to distinguish between the data that is better suited for batch processing (e.g., subscription payments, quarterly reports) and the data they need to have instantaneously (e.g., customer behavior).

Additionally, businesses should focus on providing non-technical teams with timely access to the data they need to stave off delays.. For example, Segment offers **an intuitive, no-code audience builder** so marketers can quickly spin up personalized campaigns in minutes – no waiting on manual audience exports or SQL required.



Examples of batch processing and real-time data

Batch processing	Real-time data
Employee payroll	Product recommendations
Quarterly business reports	Customer communications
Credit card statements	Weather alerts
Subscription billing cycles	Crisis response
Inventory processing	Fraud detection

Not complying with privacy regulations

Data privacy is the practice of keeping information – especially someone's personal information – confidential. It limits who can access, transfer, or share this data without the owner's permission.

Data privacy is both a legal and ethical obligation for businesses, with notable laws including the **Health Insurance Portability and Accountability Act (HIPAA)**, which instituted strict protections around an individual's medical data in the U.S.

Then there's the **Global Data Protection Regulations (GDPR)**, which provides EU citizens with uniform privacy rights. (Even though the GDPR is EU-specific, it applies to any country that does business with EU residents.)

Staying compliant with relevant privacy legislations can be a challenge for businesses. For one, it's a fast-changing landscape, and in certain instances, it's dismantling practices that have been around for decades (re: the use of third-party cookies in advertising).

That's why it's important for businesses to embrace a "privacy by design" approach, which is when privacy controls and considerations are integrated into systems from the start.

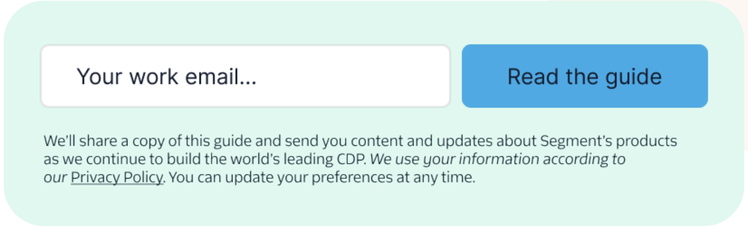
This is **particularly important in the age of AI**. As more organizations embrace machine learning and generative AI, it's essential that they're using compliant, consented data to train these models.

How to fix it

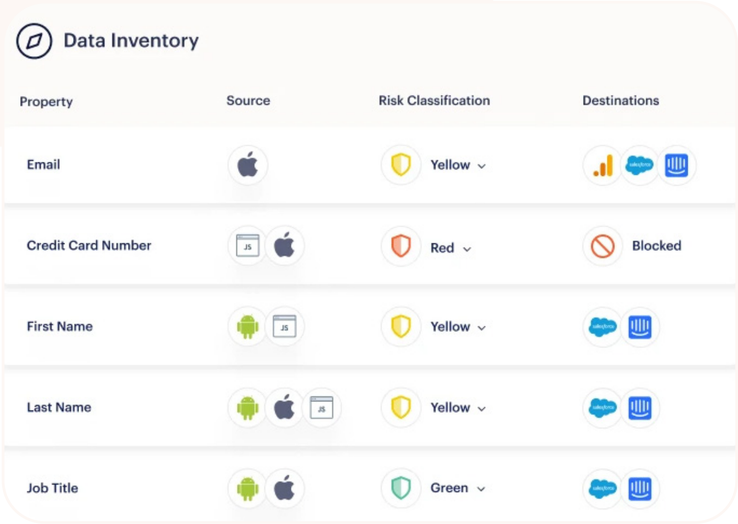
One of the fundamental steps a business can take to ensure compliance with privacy regulations is to prioritize first-party data. Unlike third-party data that's acquired from external sources, **first-party data** is gathered through direct interactions a person has with your brand (like transaction history, website analytics, and customer support interactions).

Companies that buy and sell third-party data often have no relationship with the customers whose information is being used. This makes it difficult, if not impossible, to verify third-party data's accuracy and compliance. (Not to mention, third-party data is less valuable because anyone can have it – you and your competitors could be using the exact same data to run your marketing campaigns.)

Another important step for regulatory compliance is to establish a **data privacy policy**. This is a legal document that details how a website visitor's personal data may be collected and used (and it should be easily accessible across your site).



Organizations also need to protect personally identifiable information (PII). We recommend using the **principle of least privilege** to do this – with Segment, we have a feature called **PII Access** that helps companies implement fine-grained controls over who internally has access to sensitive customer information. We can also automatically detect and classify personally identifiable information based on its risk level (e.g., someone's social security number would be high risk).



Then, there's consent management, or having the ability to honor user suppression and deletion requests at scale. You can learn more about Segment's open-source consent manager **here**.

Chapter 03

Protect data quality at scale with Twilio Segment

YOUR DATA GOVERNANCE CHECKLIST:

Capability	Description
Tracking plan	Use a customizable tracking plan or spec to align your business teams on which events you collect, what they mean, and what business metrics they drive.
Data validation reporting	Automatically test your data against your implementation spec or tracking plan to identify any data that doesn't match.
Data violation reporting	Drill into specific data quality issues, including both data type and data source, for actionable context and quick debugging.
Notifications and alerts	Receive realtime notifications and alerts when your data doesn't map to the spec you've outlined.
Enforcement settings	Block, disable, or quarantine unwanted events without making changes to your implementation.
Cross-platform standardization	Apply standards to multiple data sources for consistent implementation across your website, apps, and servers.
Privacy consent management	Honor consent and user preferences, like deletion and suppression, in alignment with regulations like the GDPR.

Learn more about how Twilio Segment makes your data more trustworthy and actionable, **schedule a demo**.

As you can see, the qualities that can make data "good" or "bad" are often interconnected. For organizations to ensure high-quality data, they need to begin with good data governance – or establish a set of policies and best practices that dictate how data will be collected, unified, stored, activated, and deleted. Of course, as businesses scale, data governance can be difficult to uphold. That's why businesses should embrace automation when possible.

With Twilio Segment, organizations can leverage pre-built integrations (with **Connections**), or automatically enforce their tracking plan with **Protocols**. This helps dismantle data silos and proactively block bad data before it hits downstream destinations.

Segment also ensures that data is encrypted at rest and in transit. Our **Privacy Portal** helps companies comply with global privacy regulations, from being a HIPAA-eligible platform to offering regional infrastructure in the EU to comply with data residency requirements.

But equally important to collecting good data is being able to act on it. **Unify** merges real-time customer data across each platform and channel you use into a unified profile. These profiles can then be synced to your warehouse for further enrichment. Empowered by this holistic view of customers, marketers can then leverage this data to create highly personalized audience segments and to orchestrate cross-channel **journeys**.

Today's leading companies trust Twilio's Customer Engagement Platform (CEP) to build direct, personalized relationships with their customers everywhere in the world. Twilio enables companies to use their communications and data to add intelligence and security to every step of the customer journey, from sales to marketing to growth, customer service and many more engagement use cases in a flexible, programmatic way. Across 180 countries, millions of developers and hundreds of thousands of businesses use Twilio to create magical experiences for their customers.
For more information about Twilio (NYSE: TWLO), visit: www.twilio.com.