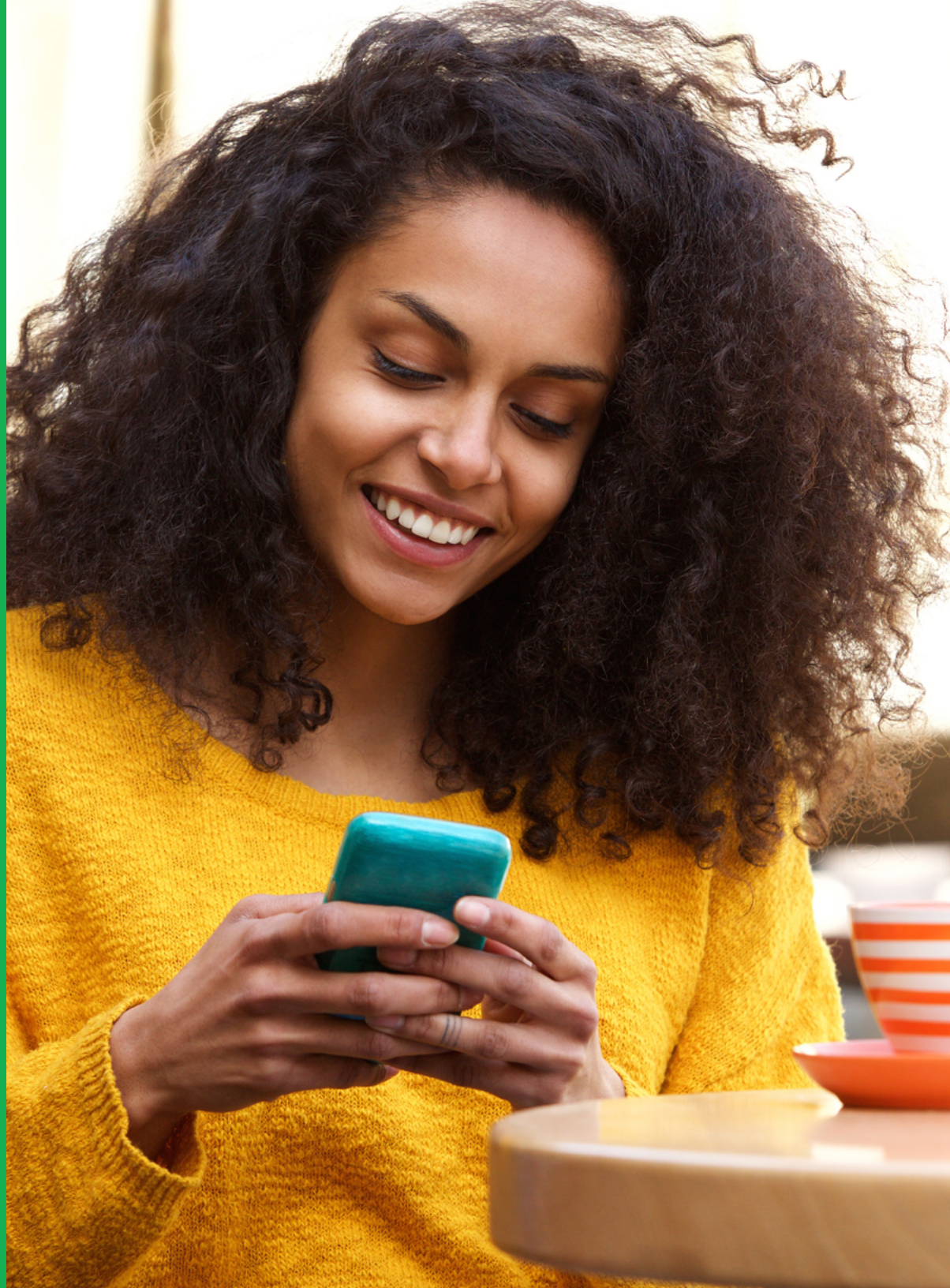
Building a Single View of the Customer With Identity Resolution

TWILIO Segment



Contents

Introduction

Chances are you've probably heard the concept of the "single view of the customer" thrown around quite a bit.

Also known as the "360-degree view of the customer", it promises to give businesses a complete view of their customers, by aggregating data from multiple different touchpoints.

However, this "single view" is usually an empty rallying cry for businesses; in their day-to-day operations, most struggle to get a 90-degree view of their customers' behavior.

This is because gathering all the data about your customers and merging it into a single record is still a largely unsolved challenge.

The crux of the issue is this–for a business to get that elusive single view of the customer, you need to have good identity resolution.

But what exactly is identity resolution, and how can it be implemented at your organization? Read on to find out.

What is **identity resolution?**

Identity Resolution, also known as ID Resolution, entity resolution, identity mapping, or record linkage, is the practice of creating a single customer profile for every customer by unifying different data sets pulled from a variety of locations, including CRM, marketing and support tools, SMS records, third party databases, and more. Data teams use Identity Resolution to connect real-world data concerning a single person from a variety of sources so that all customer data and behavior is in the same record.

Building a scalable, sustainable model for identity resolution is not a trivial task and requires extensive work from the data team. Not only do teams need a consistent approach for creating and resolving customer profiles, but those profiles need to be maintained, which is a continuous process of cleaning, merging, enriching, and porting data back and forth between system application layers, the warehouse, and a BI solution of choice. And since the average data scientist is still spending almost 40% of their time on data prep and cleaning. According to a 2021 Anaconda study, introducing another model that requires constant heavy lifting to keep it on track is a recipe for disaster.

Many intrepid data teams have embarked on an identity resolution journey, only to find it's tougher to build than it looks. However, it is possible, if you know the challenges and best practices.

Common identity resolution challenges

Identity graphs and data pipelines are complex

Identity resolution is tough, and it sometimes takes years to create a solution that you can fully trust. Sadly, many optimistic data science teams get to the final stage of launching their identity resolution model, only to discover that it is immediately broken. The challenges of maintenance are compounded as business teams continue to onboard new engagement solutions with completely different data structures, which render the approach to merging customer data into the profile out of date.

Customer data is often incomplete and inaccurate

For any given customer profile, the following inaccuracies and redundancies are common:

- The same person may submit a form with multiple email addresses and therefore have multiple profiles
- Due to typos, the same person may be in the system a variety of times using a different spelling of their name or company names
- Third party data, which is less trust worthy, may be used to create or append a customer profile
- Abbreviations, varying uses of punctuation, and omitted fields create errors

These data gaps require data teams to model a solution for handling, merging, and rejecting data as it is created. This is not a straight forward task since there are infinite error types that can be introduced to the data. We will expand on best practices for handling this challenge later.

Changing rules and regulations threaten compliance

Both internet providers and governments are making moves to ensure personalization doesn't happen at the expense of customer privacy. So your solution for identity resolution must strike the right balance of personalization and privacy, even as compliance rules continue to change. For example, Google Chrome, the world's most popular browser, has promised all third-party cookies will be deprecated by 2024, while fellow industry giants Apple and Mozilla have taken similar steps.

Meanwhile, governments have introduced GDPR, CCPA, and industry-specific privacy regulations like HIPAA for healthcare, all making customer data collection, and subsequently personalization, more difficult.

Identity graphs are difficult to maintain over time

In 2022, Chief Martech identified over 9,900 marketing technology solutions available to businesses. Your business teams may not use all 9,900, but they probably use more than they can count on two hands. With potentially millions of data points coming into the business on a weekly – or even daily – basis, there are bound to be challenges capturing and recognizing data from all 3rd party sources, especially when there are conflicting inputs, or difficulties tracking the historical/ changes to profiles.

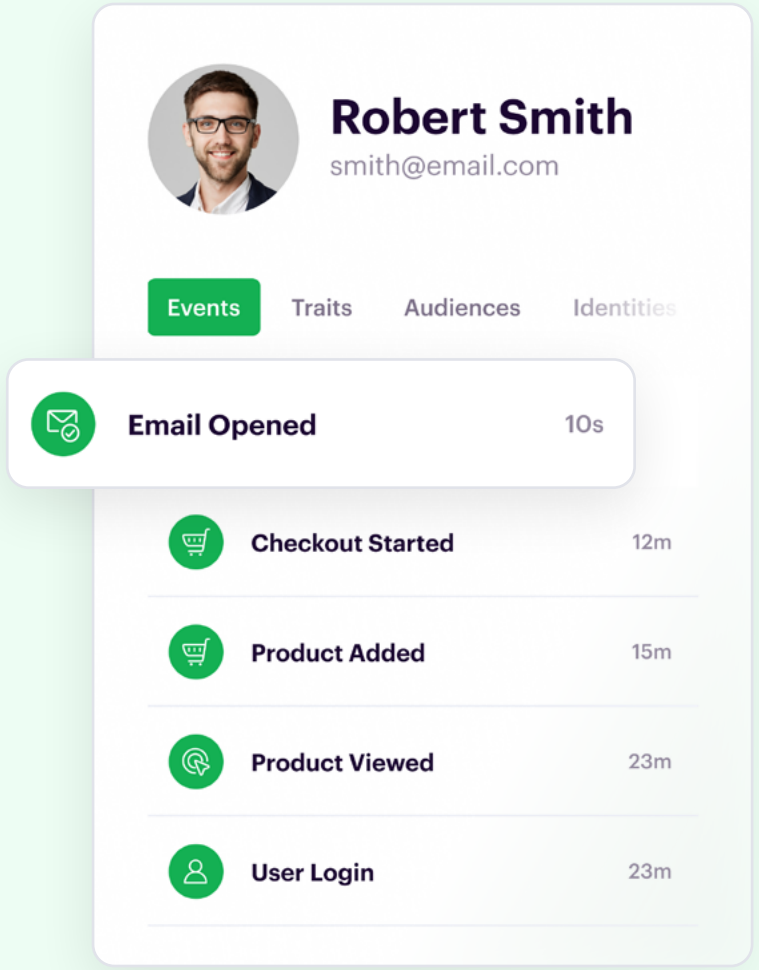
Without a continuous process for capturing, cleaning, and merging new data inputs, it's impossible to capture a real and true snapshot, let alone maintain an accurate profile over time.

Resolved customer data is stuck in the warehouse

While there is growing interest in using the warehouse as the source of truth for identity resolution, there are challenges to this approach.

Data comes into the warehouse from a variety of sources and a variety of formats, often with different schemas, and manual-based solutions for turning raw data into clean customer profiles can be incredibly complex and time consuming to manage. Additionally, creating pipelines for the multitude of data inputs and assigning meaning and relevance to data that is often incomplete, inaccurate, and in a variety of formats, is a huge challenge, even for the most qualified SQL expert.

The goal of identity resolution is to help the business understand the full picture of the customer journey so it can become accessible and actionable for business users through analytics. Yet, many parts of the business will often never see the results of these models outside of an analytics dashboard within a BI tool - if they ever see it at all. This makes running self-service queries, creating audience segments, or actioning complex use cases like marketing attribution difficult if this data isn't available in the tools teams already use today.



Getting started with identity resolution: **best practices and considerations**

With changing guidelines around customer privacy, exploding customer touchpoints, and a myriad of data challenges to solve, it's no surprise that identity resolution, the touchstone of personalized marketing, remains a critical challenge. But it is possible to develop an effective solution for Identity Resolution with best practices at the core.

Customer Data Platforms, like Segment, help businesses collect, clean, and control their customer data. Segment includes identity resolution, the ability to merge customer activity into a single profile in real time.

Our solution – and approach – to identity resolution was developed through our experiences with over 25,000 customers, across a wide variety of industries. Whether you work with a CDP provider, or attempt to develop your own identity resolution pipelines (more on that later), here are some considerations and best practices we've learned that can simplify the complex and challenging world of identity management.

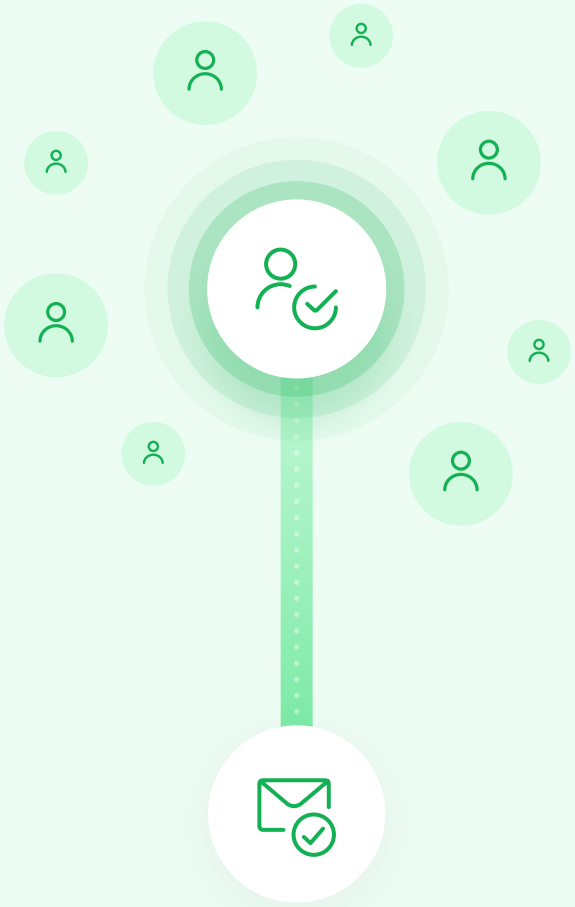
Start with a true universal identifier

One of the first challenges most data science teams face when approaching identity resolution is that (some) of the business teams may believe the problem has already been solved. This is because many off-the-shelf customer communication tools treat “identity” as matching on an email address or a hashed email address. This approach works for vendors, but falls flat for businesses as email is not a true universal identifier. If a team leveraging “email” as an identifier attempted to sync every single one of your sources – server, mobile, cloud, web, etc. – on behalf of a user, it would result in a bunch of disconnected tables that would need to be stitched together.

Consider a user with an anonymous ID AABBCC who is then de-anonymized to become user 1243. In this scenario, every event generated before that customer logged in will be stored in your warehouse, forever, with just that anonymous ID. If you want to activate that data and understand user 1243’s whole history with you, you’ll have to look that anonymous ID up. Your data will certainly be fragmented with just an email, just an anonymous ID, or any one of a number of identifiers, leaving you with the work of resolving, each and every time, who that identifier actually represents.

True identity solutions should connect data to consumers in a privacy-compliant way by establishing brand-unique identity resolution rules that create a single view of the consumer across all channels.

This is where CDP solutions like Segment can offload the heavy lifting by providing a canonical identifier that can be the point of consistency that all other identifiers are merged with. Starting with externalIDs, which are identifiers pulled into your CDP from an external data source, like user_id, Android and iOS IDs, Google Analytics, anonymous_id, and group_id., you can match external identifiers to a canonical_id, which is then the source of truth as the profile is built over time.



Simplify the approach of building your data pipelines

Building and maintaining the data pipelines required for effective identity resolution typically consists of following steps:

1. Collecting data from sources
2. Processing/transforming data in a usable format
3. Activating the data in destinations used by partners teams
4. Orchestrating steps 1-3

As more business applications, customer identifiers, and dependent teams are added, these steps become increasingly difficult, and often have to be reimagined as the business changes. This is where solutions like Segment can provide a lot of value by streamlining these steps, even as the business goes through change.

With Segment, data teams can easily collect data from a website, server library, mobile SDK, or cloud application, and then process the raw data to build an identity graph that allows data teams to understand user interactions across web, mobile, server, and third-party partner touch-points in real time.

Customers can access the rich identity graph tables in their warehouse using Profiles Sync or build personalized experiences by calling Profile API. Additionally customers can use Reverse ETL to activate data in destinations used by their marketing and sales partners. With Protocols, customers can ensure that data adheres to data collection best practices, thereby making it simple for data teams to focus on solving business use cases rather than fixing data problems.

Take a more fixed and flexible approach to data merge rules

Identity resolution is a complex process that's tied to a businesses unique approach to customer engagement. So it makes sense that every identity resolution solution is just a little bit different, and it makes even more sense that every data scientist worth their degree wants to put their stamp on a solution that is tailored for your individual business. While it's true that flexibility is needed to handle the complex realities of identity resolution, there is such a thing as too much flexibility. Because the behavior of overly flexible systems inherently can't be predicted or controlled, it is prone to risk. It happens all the time – a marketer notices a discrepancy in the identity graph, and the data scientist who built the thing can't satisfactorily explain how it emerged or how to prevent it. When that happens enough, the marketer wants to move on from the solution.

The best approach builds rigidity around what should be fixed, so data science teams can focus on the highest impact areas for the business, and a deterministic approach (more on that later) to identity resolution solves this challenge neatly for data science teams.

Segment's approach to identity resolution enables you to provide many identifiers for the same person (such as user_id, email, phone, device_id, anonymous_id), and then set the priority of matching to control how profiles are stitched together. It's a powerful model which has helped us create over 5B unified customer profiles from billions of stitched together events. But this model requires *alignment* and *consistency* in the way you stamp values for each identifier onto your events. Otherwise identity resolution won't work.

In order to achieve an aligned and consistent approach to identity resolution, Segment relies on a 100% deterministic model, based on first-party data. Our approach is called "deterministic" because it requires exact matches (instead of "fuzzy" or "probabilistic" matches) on identifier values to unify events into a single profile. This means identities are resolved based on what you know to be true as opposed to resolving identities based on what you predict to be true (probabilistic identity resolution.) At Segment, we believe that deterministic is the best approach for identity resolution because it's based on first-party data your customers actually produce.

While Segment's deterministic identity resolution might seem overly rigorous, it's actually highly beneficial. It enables 100% reliable profile unification, and it honors the exact first-party data a user provides to you, so your rationale for merging profiles (or keeping them separate) is completely transparent.

While Segment has always taken the approach that it's better to be ridged in some areas and provide flexibility in others, so you have a consistent process for categorizing and processing data inputs across all your applications, increasingly rigid privacy laws have further supported the deterministic approach. Global organizations are most likely aware that the EUs General Data Protection Regulation (GDPR) already prohibits companies from taking advantage of probabilistic identity resolution. While the GDPR may seem like the most strict form of privacy regulations today, countries and technology companies around the world are taking action to be more respectful of consumer privacy. For example, Apple's latest iOS 14 release includes a new privacy feature that gives users more control over which apps can track them across sites and apps for advertising. As operating systems and browsers release more privacy features, and more users choose to opt out of tracking, the value of a probabilistic strategy will continue to decrease.

Understand the role of the warehouse

Innovative warehouse providers like Snowflake, BigQuery, and Redshift have enabled data teams to execute complex queries and transformations more quickly with less fear of resource constraints. Consequently, some teams have attempted to use the warehouse as the business source of truth, and build SQL models for computing key business metrics such as LTV or health score with data straight from the warehouse.

While the separation of storage and compute model has made the data warehouse a much more resource-friendly place to store all customer data, attempting to perform identity resolution in the warehouse still has its drawbacks. Ideally, the warehouse would come equipped with a sophisticated ability to understand how to match various records to represent one customer against a set of criteria, but this is not the case. Most data teams that try to manage an identity graph with more than two or three customer identifiers using SQL find very quickly that it is intractably time consuming, expensive, and complex.

Solutions have existed on the market to sync data from their data warehouse to their customer-facing teams' tools, but this is a simplistic look at what is actually required to create a unified profile with a canonical identifier that can be leveraged across all your business applications to execute across the complex use cases that move the business forward.

In short, while the data warehouse can be a source of truth- it is a hammer that can make every problem look like a nail if we aren't careful.

Make portability a core component of your identity resolution strategy

The beauty of identity resolution is that your whole business is in sync, you aren't having to provide the identity graph to everyone and make it work for everyone. You have a solution for keeping profiles up-to-date via APIs, and can allow teams like marketing to build traits and campaigns in the solutions they use, while data science teams can build models in the back end to answer the businesses' toughest questions. You can even use APIs to move, clean, to pull in demographic, behavioral, financial, lifestyle, purchase, and other data compiled or licensed from third-party sources, such as online news sites, purchase transactions, surveys, email service providers (ESPs), motor vehicle records, voter registration, and other public records.

Not only do solutions like Segment allow data to be incorporated into the profile via APIs, but once the profiles are built, they can be easily moved from the warehouse to all business applications, and back again. With the newly launched Profiles Sync, Segment's identity-resolved customer profiles are portable. Data teams can sync customer profiles to the data warehouse, and create valuable audiences in the data warehouse that power personalized customer experiences. With Reverse ETL, data teams can easily activate enriched profiles, and other valuable data stored in the data warehouse.

With Profile Sync and Reverse ETL as part of your Identity Resolution solution, an up-to-date, golden profile can be synced easily with both the warehouse, and the business applications that rely on up-to-date data. All the basic maintenance and cleaning work is handled, so data science can focus on the more complex use cases, and you don't need to buy another Reverse ETL tool to create more fragmentation and silos.

Beware of the costs of building vs buying

While many teams go the DIY identity resolution route to save money by reducing the involvement of third party vendors, the reality is that DIY identity resolution requires the support of a lot of vendors!

It is possible to create a basic version of an identity graph using SQL and data in the warehouse, but making the benefits of identity resolution available to the business will require investment in a host of solutions to make the data portable. Once you have resolved customer profiles, you need to make that information available to business teams in their systems of choice so they can deliver personalized messages, campaigns, and more. This information needs to be actionable, and as more customer interactions change the customer record, this information has to come back into the system of record, and update the customer source of truth.

The more underlying infrastructure (servers/VMs, compute, storage) and vendors (like AWS) that need to be connected to purpose-built tech stacks (like martech, CRM, support, and ecommerce tools), the more complicated the data pipelines become, and the more data tools you bring into the mix, the more vendors you need to work with as well.

A classic configuration can require up to 5 different tools and 5 different vendors:

- Event Streaming
- ETL
- Reverse ETL
- Data warehouse
- Data transformation

As the requirements to connect more and more systems grows, so does the time-suck and expense of navigating the myriad of protocols, languages, interfaces, and UI/UX choices in an ecosystem of providers. Many of these tools lack “out of the box” connectors and so connecting the dots between systems requires a lot of custom code, especially for legacy systems that require a different approach than more agile cloud based systems and lack the API coverage to easily connect to new tools.

This requirement to evaluate, purchase, setup, and maintain new solutions to make identity resolution outputs useful to the business dramatically increases the costs (time and money) of DIY identity resolution, and makes many data teams regret the day they decided to pursue building a “simple” in-house solution.

Case study: **Fender and Twilio Segment**

USE CASE
Customer 360
Customer data infrastructure
Data stack modernization
Data powered performance marketing
Data powered digital customer experiences
Customer insights and analytics

INDUSTRY
Consumer Goods and Services

HEADQUARTERS
Los Angeles, California

REGION
NAMER

PRODUCTS
Connections, Protocols, Twilio Engage

How Fender uses identity resolution to increase customer engagement and lifetime value

Founded in 1946, Fender's guitars, amps, pedals, and accessories have been played on more stages than any other company in the industry. With more than 75 years in the business, the Southern California company has established a worldwide influence that extends from the studio to the stage to the at-home consumer. Fender worked with Segment to increase activation for Fender Play, its music lessons app, and graduate app subscribers from free to paid accounts.

At the core of Segment Unify is Identity Resolution. Fender uses Identity Resolution to merge the history of every customer into a single profile so they can personalize a Fender Play user's learning experience and display the right product to that e-commerce customer.

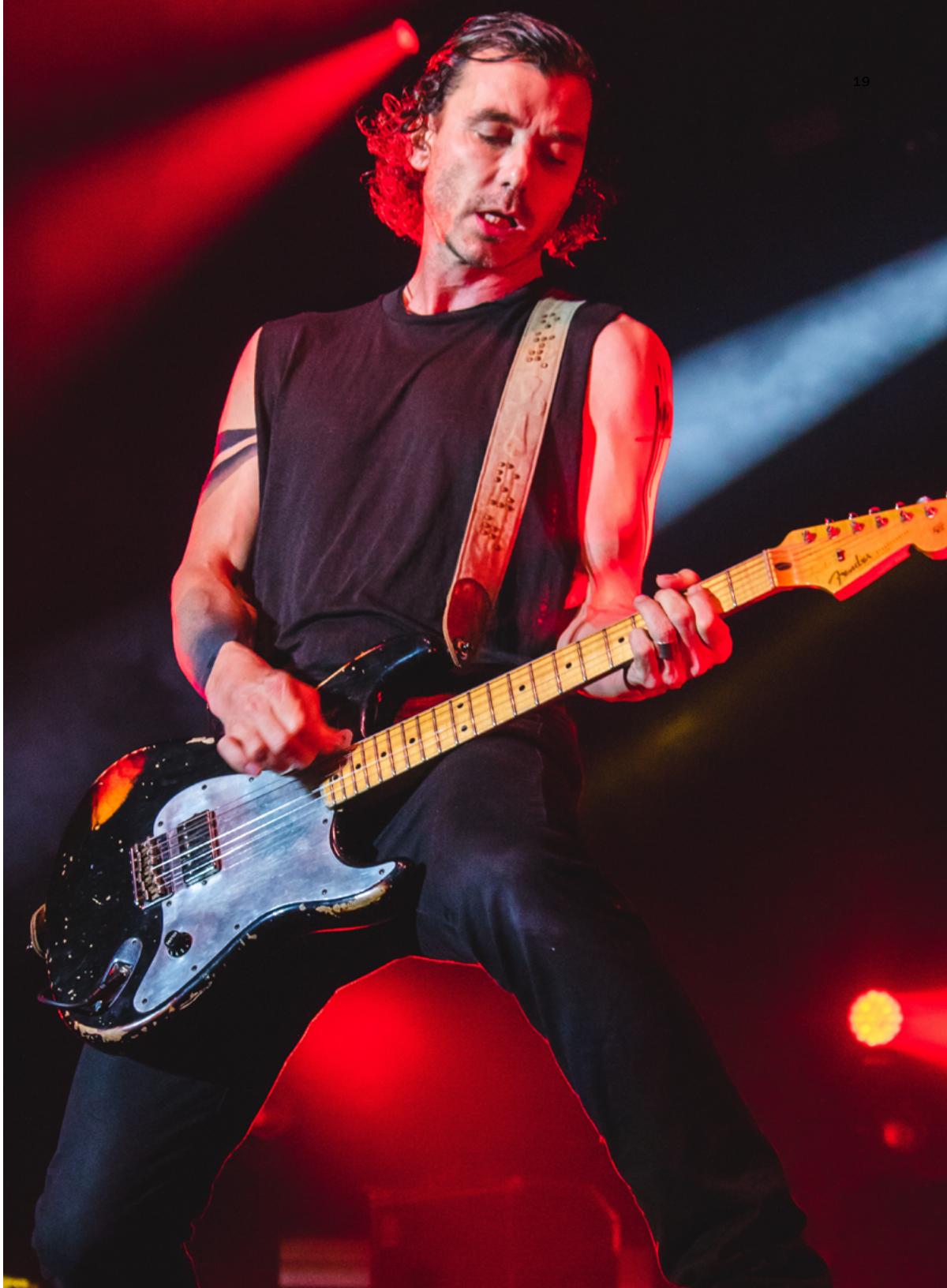
The power of Identity Resolution is that it allows Fender to know that Johnny B. who browsed guitars on mobile and John Goode, who viewed amplifiers on its website, is the same johnny.b.goode86 who has liked every single image of ukuleles that Fender posted on its Instagram account.

Tracking user activity across multiple touchpoints ensures Fender never misses a beat. With access to consolidated customer data, creating audiences with similar user traits is easy. And it lets the marketing team tailor campaigns with product recommendations and offers that probably has some customers thinking about getting the band back together.

With identity-resolved customer profiles Fender personalized offers and lesson plans in Fender Play and Increased active paying users by 5%.

“Segment Unify is really exciting because it’s the future of Fender. Educating ourselves about our users in a responsible way – to give them the ideal learning and shopping experience – is what’s going to help them become players for life.”

Claire Armstrong, Director of Digital Product Management, Fender



Case study: **Sanofi and Twilio Segment**

USE CASE
Data enabled sales and support teams
Data powered performance marketing
Customer insights and analytics
Customer 360

INDUSTRY
Healthcare

HEADQUARTERS
Paris, France

REGION
EMEA, APAC, NAMER

PRODUCTS
Twilio Engage, Protocols, Connections, Unify

sanofi

How Sanofi builds “golden profiles” to increase omnichannel engagement and improve patient outcomes

Sanofi, a leading global healthcare company, provides life-changing treatments, medications, and vaccines to millions around the world. Fueled by data and digital technologies, cutting-edge science and manufacturing, Sanofi aims to transform the practice of medicine through the creation of innovative treatments for patients across the healthcare spectrum. Sanofi helps healthcare providers (HCPs) such as doctors, hospitals, and pharmacists provide the best possible care to patients by continually educating them on medications that fit patients’ specific needs.

Sales representatives at Sanofi have regular on-site contact with HCPs, but Sanofi struggled to connect these customers’ digital interactions with offline channels. This created gaps in Sanofi’s understanding of the entire customer journey and missed opportunities to connect online and offline customer touch points. Sanofi couldn’t optimize distribution of information to customers about new medications and treatment plans because It lacked the 360-degree customer view it needed to enable personalized customer communications.

To provide a better experience for healthcare providers, Sanofi set out to collect its own first-party data and build accurate, real-time customer profiles that could power digital-first engagement and personalization.

Initially, Sanofi considered building its own data solution and ecosystem, but ultimately decided it would be inefficient to fully recreate a solution that already existed.

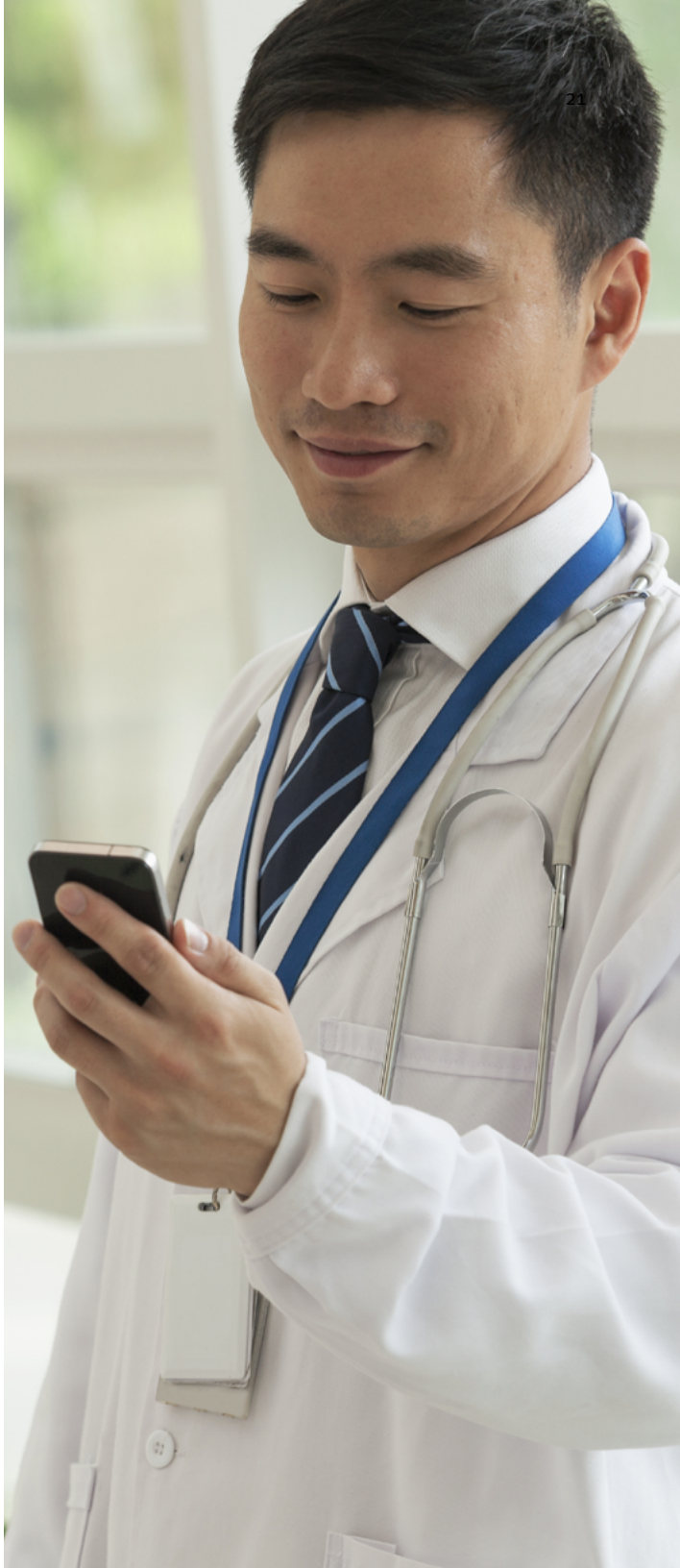
Sanofi used Segment Unify to help launch OneView; golden profiles, built using online and offline customer data from multiple sources and used to target omnichannel marketing campaigns. With OneView, Sanofi is able to better educate HCPs on medications and treatments, improving the customer experience and patient outcomes. Ultimately, the decision to make the Segment profile a node in their data graph enabled Sanofi to take ownership of its first-party data and accelerate implementation. Before Segment, it took three-plus days to activate customer data; now Sanofi is activating data in under three hours, a 95% increase in efficiency.

With Twilio Segment, Sanofi can:

- Collect and centralize data from multiple sources and activate it immediately, enabling better engagement with healthcare professionals.
- Synthesize raw data into real-time, identity-resolved customer profiles
- Deliver complete customer profiles to the warehouse with Profiles Sync, enabling profile enrichment and valuable audience insights.
- Use Reverse ETL to easily send profile data back out of the warehouse to any tool in their tech stack for laser-focused targeting and personalization.
- Improve healthcare providers' understanding of new therapeutic areas, leading to better experiences for both doctors and patients.

"Previously, we would work with our data by moving it, then using it. Now, with Segment, we can use data while we move it. The big benefit of Segment is not only its ETL capabilities to move data from point A to point B, but Segment is also able to unify data quickly in the warehouse before we send the data to campaign tools."

Rick Troiani, Director of Omnichannel Engagement Architecture at Sanofi



About Twilio Segment

Twilio Segment is the leading CDP with more than 450 pre-built integrations to different data sources and destinations. It provides a complete solution that eliminates the need for manual data cleansing, complex data engineering processes, and analytics reporting functions.

By automating all of the backend customer data operations, Twilio Segment puts companies in a position to get the most out of their first-party data and retain customers at a higher rate.

As consumer sentiment, industry trends, and regulatory enforcement push companies away from depending on third-party data, the need for an alternative source of customer data cannot be understated. First-party data is the solution, bringing a competitive advantage as it fills the gaps where third-party data falls short: accuracy, relevance, and building customer trust.

Schedule a demo to learn how to get the most out of your customer data with Twilio Segment.

Recommended reading

The Customer Data Platform Report 2023

Our new Customer Data Platform Report analyzes real data from the Twilio Segment platform (nearly 12 trillion API calls to be exact) to examine the biggest digital trends happening in customer engagement.

Get the report >

Digital Advertising in the Cookieless World

Learn how to deliver personalized and more transparent experiences in the absence of third-party cookies.

Download the guide >

Customer Acquisition Cost (CAC): A Guide for 2023

In this guide, we share the CAC formula and explain how to calculate it. We also look at what makes a “good” CAC, how to determine an LTV:CAC ratio, and what steps you can take to improve both metrics for long-term success.

Download the guide >

Thanks for reading

If you would like to learn more about what Twilio can do for your business, please contact the Segment sales team.