

S E T H F A M I L I A N



Customer Data Maturity



How the World's Biggest Companies Win
by Unlocking Customer Potential

Copyright @ 2023
Twilio Inc. All rights reserved.

Customer Data Maturity

How the World's Biggest Companies Win
by Unlocking Customer Potential

by

Seth Familian

Acknowledgements

This deep dive into the world of customer data wouldn't have been possible without the editorial expertise of Geoffrey Keating, Jim Young, and Kathleen Velasquez, the design talents of Schuyler MacFayden and Keenan Brand, and the visionary guidance of Steve Davito, Katrina Wong, and George Chen. It takes teamwork to make the dream work, and this team is truly a dream to work with.

Preface

"Data is the new oil" has quickly become the catchphrase of the digital era.

Data is the key to unlocking everything from differentiated customer experiences, real-time personalization, AI and innovation, and much more.

The challenge? While businesses today are collecting more data than ever before, they're also struggling to make sense of it all.

Data remains trapped in department-specific silos. Tech stacks are fed incomplete and inaccurate data. Loose data standards lead to mistrust between teams.

This has left many organizations in a frustrating position – data rich, but poor in insight. A recent, sobering study reports that just 26.5% of business leaders claim to be data-driven as an organization.

But with investments in big data and AI on the rise, there's mounting pressure for businesses to forge ahead and make good on their data investment. Investing in data collection without also focusing on how to clean, unify, and democratize said data is throwing money down the drain.

But if you were to take a peek inside the top websites and mobile apps used by millions of people today (think industry titans like Amazon, Netflix, Meta, et al.) you'd see an interesting pattern. Not only are they able to collect customer data at scale, they have the culture and technology to make sense of that data—and to act on it with speed and precision.

For the past decade, Twilio Segment has helped 25,000+ businesses, from SMBs to the Fortune 500, tap into the power of their customer data. We empower them to capture data from every customer touchpoint, understand insights, and take action anywhere, fueling an engagement flywheel that grows stronger with every cycle.

Our powerful data infrastructure unlocks the capabilities of predictive and generative AI, helping businesses better engage with customers. In turn, this process generates data that helps you predict more accurately and automate based on these insights. This cycle of data generation, prediction, and personalized engagement keeps the flywheel spinning, which accelerates revenue generation, and cultivates Customer Data Maturity.

What follows is a comprehensive framework to navigate the path to Customer Data Maturity. Every step of this journey – from foundational basics and advancing data best practices to adaptive maturity and sophisticated AI-driven customer experiences – is critical to transform your business into an organization that embodies a data-driven ethos.

This methodology has been the backbone of success for some of the world's biggest companies – e.g. Allergan, IBM, Dominos – empowering them to unlock untapped customer potential and win in their respective markets.

Throughout this book, you'll find real advice from real customers who are living and breathing customer data every day.

Supporting Resources

Throughout this book we reference a number of blogs, case studies, and digital guides. Scan the QR code below to access these resources online via segment.com/customer-data-maturity-resources



Introduction	1
01 Foundational Maturity	13
Foundational Data	15
Outcome-oriented data • Succinct data • Consistent data Engagement-focused data • Data respectful of privacy	
Foundational Profiles	27
Aligned identifiers • Compliant audiences Explicit preferences • Thoughtful spaces	
Foundational Engagement	41
Single channel • Single step • Timed out	
Assessing Foundational Success	46
02 Advancing Maturity	51
Advancing Data	53
Audited • Consolidated • Monitored • Replayed	
Advancing Profiles	67
De-anonymized • Enriched • Synced profiles Inferred preferences	
Advancing Engagement	77
Multi-step • Channel optimized • Context aware	
Assessing Advancing Success	85
03 Adaptive Maturity	91
Adaptive Data	93
Revalidated • Scaled governance • Automated docs Scaled SOPs	
Adaptive Profiles	109
Clustered tiles • Predictive traits • Governed audiences	
Adaptive Engagement	121
Interconnected journeys • Omnichannel Centralized ROI analytics	
Assessing Adaptive Success	126
Conclusion	129

The dark age of data silos

The term “data driven” seems to be everywhere these days, describing everything from decision-making strategies to technical infrastructure to cultural initiatives. Yet behind this trendy buzzword lurks a very real struggle encountered by organizations of all sizes and industries: how to deliver compelling, personalized customer experiences while respecting privacy.

Your customers **expect** these experiences, yet many organizations struggle to meet those expectations because they’re relying on bad data, siloed tools, and fragmented engagement. An organization might invest in a few dozen tools for marketing automation, CRM/helpdesk, and analytics across the stack, but any data that might be found in these disparate sources are siloed from one another.

When these tools aren’t integrated together, the organization as a whole is prevented from gaining a complete view of the customer. The partial view of the customer makes it impossible to deliver personalized experiences at scale, hindering your ability to attract or retain customers against your competitors.

Furthermore, bad data makes it difficult to assess, say, how to target the right audience on the right channels when building budget-conscious marketing campaigns. Marketing budgets are wasted targeting the wrong people because of bad data, and engineers spend most of their time on ETLs (Extract, Transform, and Load) rather than improving the core product.

We call this state of disorganization ***data immaturity***. Failing to overcome the acute pain of disorganized data means being at a competitive disadvantage to those market players who invest the resources to deliver what customers want.

From silos to revenues

In reading this book, presumably your goal is to develop data strategies that actively unlock business opportunities. The hope is that eliminating these data silos will enable your business to deliver engaging, personalized customer experiences at scale, bolstering retention and cross-sell opportunities.

The path to data maturity first and foremost requires adopting a **CDP (customer data platform)** such as Twilio Segment to gain a complete understanding of your customer. Your CDP serves as the essential foundation of the modern data stack by integrating and organizing data from every data source, platform, and channel to create a unified customer profile.

From there, you can activate your data to deliver personalized customer engagement to the right customers on their desired channels, which carries the added benefit of efficiently focusing your marketing budget on targeted campaigns.

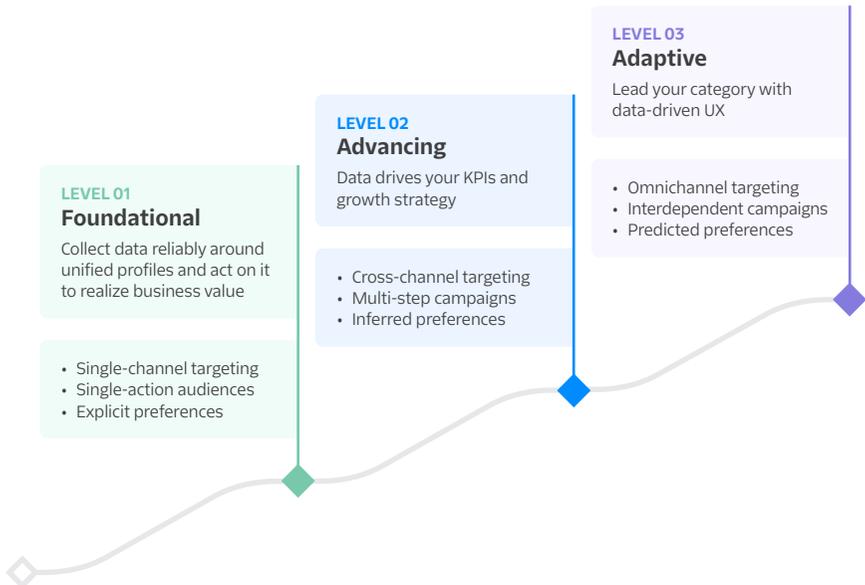
In this way, **your CDP is an essential business acceleration tool** that helps evolve organizations away from messy, siloed data and toward scaled leverage of robust customer profiles and insights which drive ROI.

The maturity journey

So, how do we get out of the rut of messy data and achieve **data maturity**? It's not an easy problem to solve, but the rest of this blueprint is dedicated to serving as your comprehensive step-by-step guide towards achieving data maturity with Twilio Segment.

Adopting a CDP alone won't instantly fix all of your problems with bad data. If the customer data fueling your CDP isn't refined, then it'll greatly hinder the performance of even the most sophisticated and powerful CDP. We'll focus on how to turn your customer data into the fuel that ignites business acceleration with your CDP.

Building effective customer engagement is itself a journey – one which requires work, discipline, and relentless focus. This journey, we've found after guiding thousands of CDP implementations, follows three distinct levels:



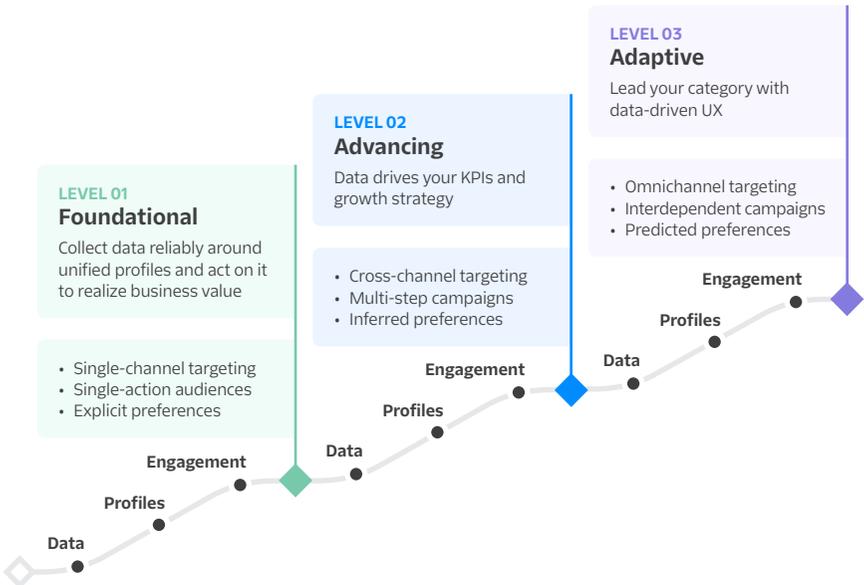
Each level above works towards a substantial customer engagement maturity milestone which drives tangible business value; even Foundational maturity can have a strong impact on your organization’s operational efficiency, revenue generation, cost savings, and risk mitigation.

The three pillars

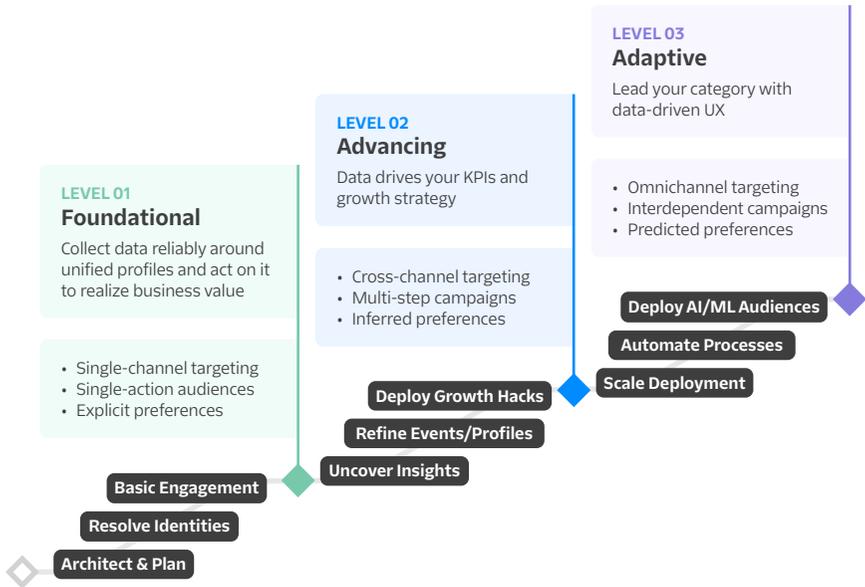
Within each level of maturity, there are three core “pillars”:

- 1. Data:** what should be ingested to achieve meaningful engagement?
- 2. Profiles:** how does that data resolve into increasingly complex profiles?
- 3. Engagement:** how will data and profiles be leveraged to achieve a business outcome?

These focus areas repeat across each maturity level:



Together, the pillars, levels, and outcomes comprise our Customer Data Platform (or CDP) Maturity Framework. **This framework is designed to be executed in order**, and involves a series of activities to perform at each level:



When executed according to our best practices (covered in the chapters ahead), this model results in tangible outcomes which drive lasting business results. While it's tempting to leapfrog past Foundational activities into Advancing and Adaptive, we've found that **the most successful and sustainably mature Segment customers first establish solid Foundational maturity before advancing into later levels** of the journey. Otherwise, you'll waste precious time and engineering cycles reworking your initially captured data and profiles to achieve more advanced use cases.

Leveling up

Transitioning to the next level is no easy feat, but we've worked with thousands of businesses to help them move along the data maturity curve. And that's exactly what this playbook is intended to help you achieve too. In the pages that follow you'll learn about the key building blocks of each Level and Pillar of the maturity model – from Foundational Data, which focuses on consistency and privacy, to Adaptive Engagement, which leverages AI and ML models.

Here's a high-level look at the topics we'll cover across each maturity level and pillar:

Operational Maturity

Through outcome-oriented activities

	Data	Profiles	Engagement
Foundational	<input type="checkbox"/> Outcome-oriented <input type="checkbox"/> Succinct & consistent <input type="checkbox"/> Engagement-focused <input type="checkbox"/> Respectful of privacy	<input type="checkbox"/> Consistent <input type="checkbox"/> Compliant <input type="checkbox"/> Explicit <input type="checkbox"/> Thoughtful	<input type="checkbox"/> Single-channel <input type="checkbox"/> Single-step <input type="checkbox"/> Timed-out
Advancing	<input type="checkbox"/> Audited <input type="checkbox"/> Consolidated <input type="checkbox"/> Monitored <input type="checkbox"/> Replayed	<input type="checkbox"/> De-anonymized <input type="checkbox"/> Enriched <input type="checkbox"/> Implicit <input type="checkbox"/> Synced	<input type="checkbox"/> Cross-channel <input type="checkbox"/> Multi-step <input type="checkbox"/> Context-aware
Adaptive	<input type="checkbox"/> Revalidated <input type="checkbox"/> Governed at Scale <input type="checkbox"/> Auto-documented <input type="checkbox"/> Scaled around SOPs	<input type="checkbox"/> Tiled <input type="checkbox"/> Predictive <input type="checkbox"/> Governed	<input type="checkbox"/> Omnichannel <input type="checkbox"/> Interconnected <input type="checkbox"/> ROI-aware

On the previous page is a checklist you can use to evaluate your CDP maturity – both today and over time. So, as you read through each of the topics listed, we hope you find this playbook to be an informative tool, as well as a reference to continually revisit throughout your customer engagement maturity journey!

Planning Exercise

Evaluate your starting point on the maturity scale.

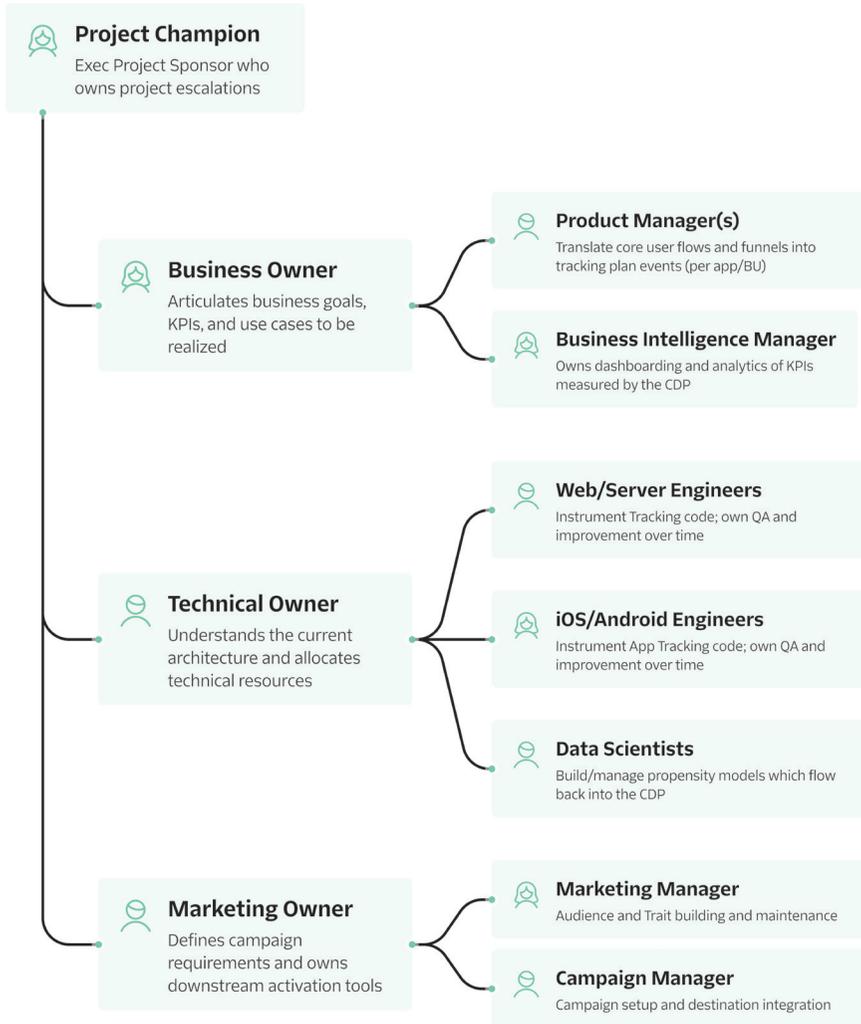
Where do you think you are today?

Even if you don't know what each word or concept means in the grid, take a stab at evaluating. You'll have a chance to revisit this evaluation throughout the rest of this playbook.

Final note: It all starts with people

Before we dive into the details of Customer Data Maturity, it's important to note that even the best technology is no substitute for getting the right people in the right seats running solid, repeatable processes. **At the center of every successful CDP implementation is a core group of stakeholders** who guide towards adoption and realization of business value. These stakeholders should cut across your organization – from Data Engineering and Analytics to Product and Marketing.

We've outlined an example model on the next page. While the number of stakeholders involved may vary between organizations, **the most important starting point is to assign a "Champion"** who will be the executive sponsor (and lead architect) for your CDP. The Champion drives alignment with Product, Engineering, and Marketing leaders to develop a unified, standard vision and data model for the entire organization.



The Champion is ultimately responsible for the implementation and activation of the CDP – often via specific Business, Technical, and Marketing owners. And part of this responsibility is to align each department around a shared vision, and ensure their accountability in upholding agreed-upon data standards and activation strategies.

Over time, one Champion can become many. But regardless of the number of Champions, their role as CDP Architect remains consistent. The Champion(s) must own (and often evolve) a centralized vision for the CDP; without that vision and proactive alignment, your CDP implementation risks losing sight of its larger business objectives.

Planning Exercise

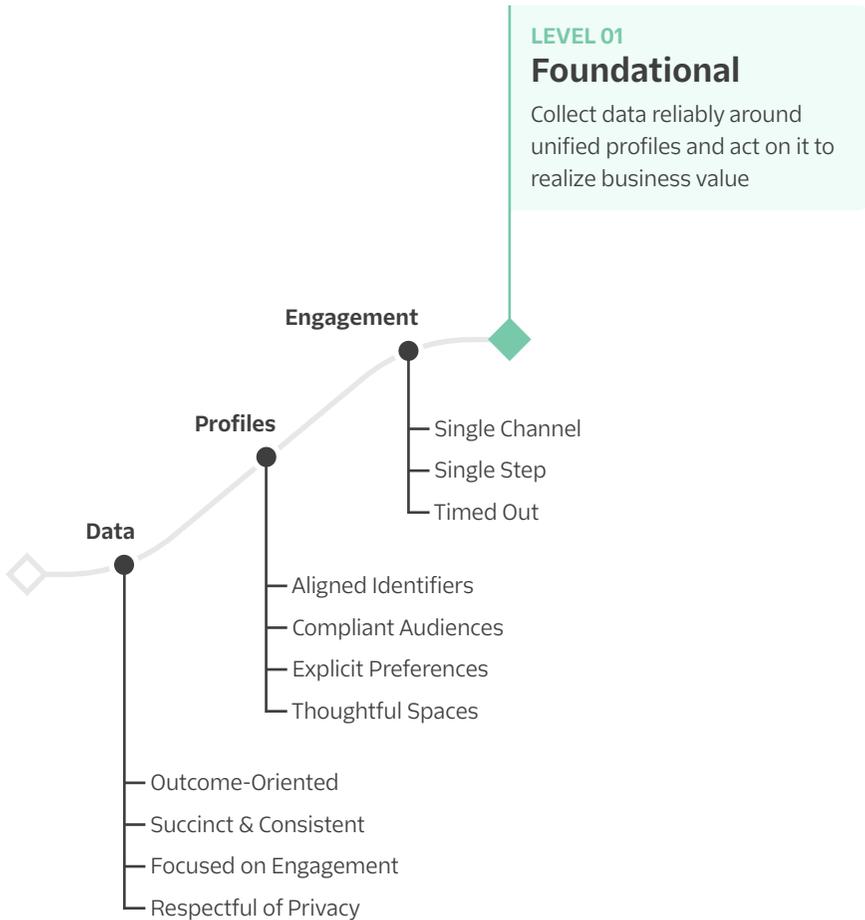
**Define your CDP stakeholders. Who is the Champion?
Who are the other key owners?**

Additional Reading

"The 4-Step Guide to CDP Success"* is a comprehensive handbook that walks through how to select the right team to deliver a successful implementation. We also recommend reading the blog post "What Does It Take to Achieve Good Enterprise Data at Scale?"* for a primer on how selecting a dream team to standardize your organization's data model results in *good data*.

* See QR code across from table of contents

01 Foundational Maturity



When building a house, you need the foundations in place before you can select paint colors or appliances that would be most visible and appreciated by the future occupants.

The same logic goes for a CDP implementation: you need to have the foundational data architecture, profiles, and engagement in place

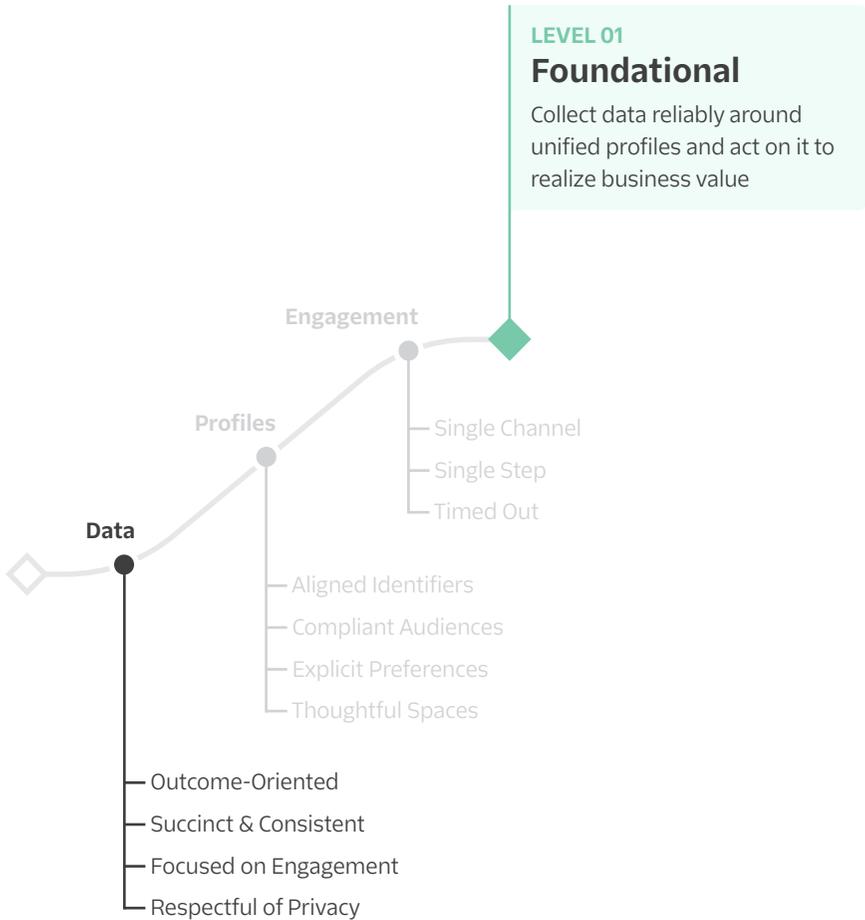
first before you can realize the more complex use cases that will create the highest value. Becoming a data-driven organization doesn't happen overnight; taking the time to lay solid foundations is a great way to set a clear path for not only how you will achieve your data objectives, but also to future-proof your infrastructure and how your CDP inevitably evolves.

In the Foundational level of CDP maturity, you will start to collect data reliably around unified profiles and act on it to realize business value. Within this phase, you will focus on engaging customers via basic tactics such as:

- **Single channels** – e.g. email, SMS, or paid advertising channels
- **Audiences of users who performed a single action** – e.g. whether they submitted a form
- **Explicit preferences** – e.g. the customer's favorite color

But before you can achieve that foundational engagement, you'll first need to build foundational data and resolve that data into foundational profiles.

Foundational Data



Building a strong data foundation requires close attention to the way you prioritize data to measure, the naming conventions for that data, the engagement signals that data contains, and alignment between your collection practices and privacy policies. In this section, we'll walk you through those core principles of generating foundational data.

Outcome-oriented data

Effective CDP planning starts with identifying the business goals your organization wants to achieve. Business goals have *measurable business impact*; they move the needle (either directly or indirectly) on KPIs like revenue, profit, and expense. Typical goals include increased average order size or frequency, increased repeat engagements, and improved signup conversion. **If you don't define these goals first, you can't plan meaningful data to track within your CDP.**

For example, if your primary business goals are to increase average order size or frequency, reduce churn, and improve signup conversion, you'll need to measure `Order Completed` (with a property for revenue), an event like `Signed In` signifying repeat engagement, and `Signup Started` plus `Signup Succeeded` to measure the signup funnel. Notice how it took just four events to measure three meaningful business goals!

If you hadn't anchored first on these business goals to define the events you'd like to measure, there's a chance you would focus first on the full detail of one particular funnel, or focus on capturing only page views across multiple BUs. And those activities would have distracted you from quick and powerful value realization.

Documenting desired business outcomes (and then anchoring your event measurement on top of those outcomes) puts you in the best position to create a data strategy that extracts maximum value from your CDP.

Outcome-anchored events can speed insight-generation, audience-building and activation – which in turn moves the needle on real business goals and revenue generation. They also help you focus on the highest-impact use cases for your CDP.

Following this approach also avoids falling prey to *vanity metrics* – high-volume data that looks impressive but doesn’t relate back to meaningful business goals or KPIs. If you’re trying to increase engagement, **App Opened** is just a vanity metric. Meanwhile, **Profile Updated** tells you someone made a meaningful change in your app after they opened it.

So when in doubt, start with the business outcome! From there you’ll be able to prioritize use cases and identify a subset of powerful events worth measuring first.

Planning Exercise

For hands-on practice, we recommend trying the “Goals & Use Cases” activity* in our Jumpstart Lite program (for existing Segment Customers on Segment University).

* See QR code across from table of contents

Additional Reading

For guidance on identifying your key goals or metrics, we suggest reading our Analytics Academy article on "Choosing Metrics that Matter". At the end of that article, you'll find additional links to content warning of the perils of Vanity Metrics.

Succinct data

Once you've identified outcome-oriented events to measure, you'll need to define a Tracking Plan – which gives you a consistent roadmap for what data is important, where it's being tracked, and why you're tracking it.

When building out your initial Tracking Plan, start with a short list of events (anchored on outcomes) and don't put pressure on yourself to measure every single data point off the bat; focus on getting the basics right!

You'll speed your time to value and create an iterative, repeatable measurement approach for your team. Remember: less is more! And oftentimes, just measuring the basics can drive very tangible value.

Let's explore an example of being succinct with your measurement. If one of your business goals is to increase new user activation, you'll likely want to measure the completion rate of your new user onboarding flow – and investigate any "falloff points" within that flow.

* See QR code across from table of contents

Let's say the flow is eight steps long. The first thing to consider is how you'll name the events at each step. Rather than naming each event uniquely (e.g. `Onboarding Funnel Step 1 Completed` , `Onboarding Funnel Step 2 Completed`), consider naming a more generic event called `Funnel Step Completed` with `funnel_name` , `step_number` , `step_name` , and `total_steps` as properties of the event. This approach cuts down on unique event names while making your funnel easier to quickly explore in analytical tools and activate in audiences.

Here's another way to be succinct in this same example: there's no need to measure all eight steps during your initial implementation! All you need to start is the first, fourth, and eighth steps – which will give you enough data to understand basic completion and falloff rates. This concept is called a “low-resolution funnel”, and it relieves pressure from your implementation team while driving faster time to insight. Once you've started gathering the low-resolution data, you can always add more interim funnel step measurement to your Tracking Plan and increase the “resolution” of the funnel to investigate additional hypotheses.

Consistent data

As you write your (succinct) Tracking Plan, you'll also want to pay careful attention to how events are named and ensure your naming conventions are consistent.

Before any API calls are sent into Segment, your CDP stakeholders should collectively align on an event naming framework that is easy to understand and ensures consistency across the board.

Segment prefers the “Noun Verb” syntax, which results in events like `Product Added` and `Order Completed` that automatically map to our downstream integrations like Google Analytics Enhanced Ecommerce. But even if you don’t go with Noun Verb syntax, what’s *most* important is that your naming convention is *consistent* across channels like web, mobile, and server. Equally important is that you’ve chosen a sufficiently abstract naming convention so it’ll scale across business units.



How to achieve consistent data in your CDP

Cross-channel analysis becomes so much easier when the same events have the same name regardless of what channel or tool you’re using. Event naming conventions also promote organizational alignment: everyone, regardless of department, can easily understand the meaning of each event being tracked. When everyone is on board, this bolsters the longevity of your data framework as a whole.

Customer Insight



Industry: B2B SaaS & Technology

IBM, one of the most iconic technology companies in the world, provides a suite of data and AI tools to help enterprises on their journey to the cloud. IBM Cloud is purpose-built with a full stack of IaaS, PaaS, SaaS, data, AI, and blockchain capabilities in support of enterprise cloud applications and AI workloads, leveraging IBM's deep domain and industry expertise.

Developing a schema that every team can use is tricky for large enterprises like IBM. After all, each product has a different function, different target persona, and different customer journeys. To get started, IBM asked its teams to identify the most important user actions that led to increased product adoption, which they referred to as “milestone” events.

IBM knew they had to strike the right balance between being general enough that data tracking could work for every team, and specific enough to differentiate between platforms – no simple task. To pull this off, IBM focused on making their product events high level (e.g. **Product Login**, **Product Logout**), and then used properties to provide important context (e.g. **Account Type**).

To do all this, IBM used Segment Protocols¹ to create tracking plans and guide relevant data from the right source to the right destinations. With Protocols, any event that didn't fit within their predetermined schema would be automatically blocked to ensure accuracy.



How IBM used Segment to improve adoption, retention, and churn

Generalized browser-side product events, with details in the properties. Super specific platform events, sent by server-side systems.

General Offering Instrumentation

- ✓ Product login
- ✓ Began process
- ✓ Finished process
- ✓ Reached milestone
- ✓ Saved work
- ✓ Shared work
- ✓ Product logout
- ✓ Page viewed

Marketing Forms

- ✓ Form completed

Trials Platform

- ✓ Trial signup

Billing Platform

- ✓ Account created
- ✓ Account updated
- ✓ Account billed
- ✓ Account deleted

Usage Platform

- ✓ User consumed
- ✓ User entitled

Marketplace

- ✓ Product searched
- ✓ Began checkout
- ✓ Finished checkout

¹ footnotes.cdmbook.com/1

Engagement-focused data

How do you track whether the users who move throughout your site are actually engaged with your products or content, even if they don't click anything? Your temptation might be to use Segment's `Page` or `Screen` call as a proxy for engagement. But this typically creates a false positive. Just because someone landed on a specific page, it doesn't necessarily mean they're engaged with the content on that page!

Segment customers solve this problem using something called “scroll depth” or “video depth” tracking. Here's an open-source repo² with some sample code. It works by first measuring the height of the page (or the length of the video) when the user lands on it, and then establishes break points at regular intervals of the page length or video (25%, 50%, 75%, 100%, for example).

As the user scrolls (or watches) past each of these breakpoints, Segment can fire off an event called `Pct Page Scrolled` (or `Pct Video Viewed`) with a property equal to the percentage depth that they've scrolled (or viewed).

This gives you valuable insights into how well specific pieces of content engage your visitors, as well as what your visitors are legitimately interested in. This second insight will prove especially powerful when we build “inferred preferences” as part of Advancing profiles, and so putting scroll/video depth tracking in place early on helps establish a truly foundational set of data we can “advance” upon in later phases of the maturity model.

² footnotes.cdmbook.com/2

Data respectful of privacy

As data privacy regulations grow increasingly complex across markets, you need to be proactive in developing a consent strategy that aligns with your company's interpretation of the law governing the capture and leverage of first-party data. It's called "first-party" because it comes directly from your customers. In contrast, third-party data is purchased from user data aggregators, who often collect data with questionable consent and share that data across different business entities without explicit permission from the end user.

At the core of this consent strategy is "wrapping" Segment's tracking libraries with a consent management tool which asks users whether and how you can track and leverage their data. That same tool then stamps those preferences on each event sent into Segment so those preferences can be respected as data flows downstream. Segment offers its own open-source consent management tool³ you can deploy on your website, and also integrates with other tools such as OneTrust and TrustArc.

It can be complex to implement this consent "wrapper" around Segment, but this work is foundational to both respecting user privacy and complying with laws and policies related to that privacy. When implemented properly, the consent wrapper achieves two things: it sets your customer's expectations around the ways their data will be used by your engagement teams, and it also establishes flags on that customer's profile in your CDP so you can respect their privacy preferences when sending data downstream.

While these wrappers help you lay the foundation for compliance with privacy laws, they do not automatically ensure compliance!

³ footnotes.cdmbook.com/3

You must combine the consent preferences you've captured with your own vigorous privacy and compliance policies to account for how those preferences will be honored in all aspects of your data leverage – from computation and audience-building to analytics and prediction. To accelerate that process, consider appointing a Data Privacy Officer to shape and enforce your data privacy policies.

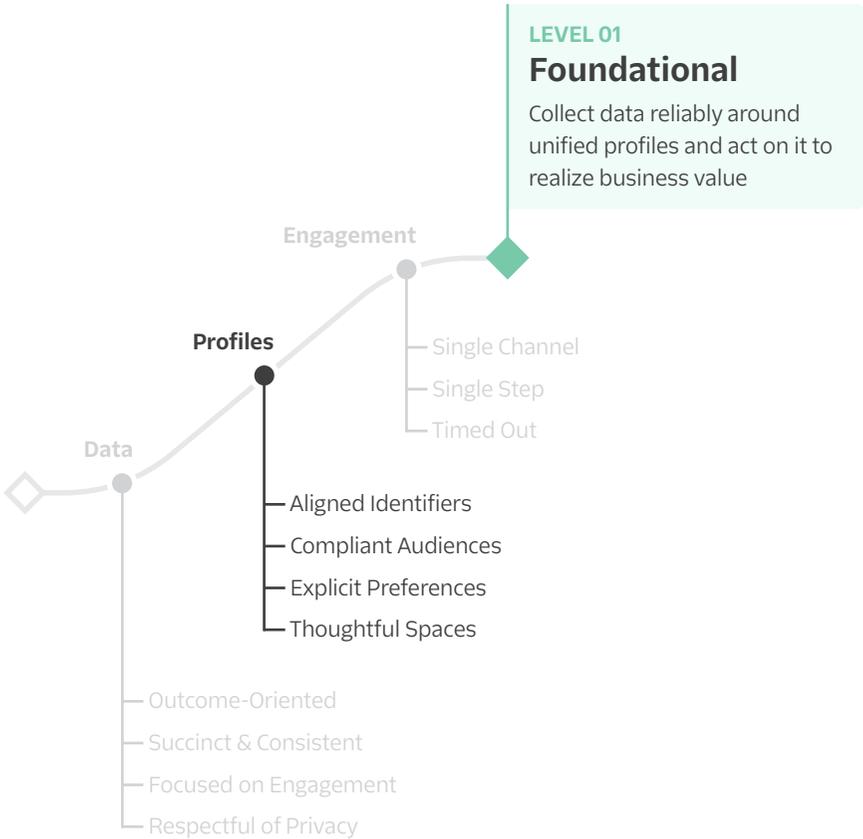
Choosing to prioritize the capture of first-party consent at the start of your CDP implementation is hard work, but it also yields a number of benefits. It helps establish trust with the people whose behavior you're measuring. It lays a solid foundation for compliance with complex laws and policies to mitigate the risk of fines and reputational damage. And it helps you ensure you're only engaging with people who actually want to be marketed to – which ends up improving return on ad spend and conversion rates.

Additional Reading

For suggestions on how to create a privacy strategy that respects user preferences, this blog post on "First-Party Data: How to Activate & Leverage Your Data"* outlines considerations when collecting and activating first-party data. To learn how to automate deletion requests, we recommend reading "Streamline Compliance with Segment's End-User Privacy Tools"*.

* See QR code across from table of contents

Foundational Profiles



Profiles unify your diverse data sources into a single view of each customer. In this section, we'll be discussing some of the critical steps you need to take to ensure that your profiles are generated and maintained successfully, so you can easily build foundational engagement.

Aligned identifiers

A major requirement for building accurate customer profiles is identity resolution: merging all of a customer's interactions across every channel and touchpoint to create a unified profile. This can seem challenging when your customers don't get identified the same way across each of your touchpoints. For example, the same person can anonymously visit your website, sign up for a newsletter, visit links from that newsletter on their mobile device, sign up for your app or service, populate their profile with additional contact information, and then call into your call center for help with a particular issue. How do you ensure that all of these activities are unified onto the same profile?

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 our customers stitch billions of events together into unified profiles. 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.

Engage

Identity Resolution

Computed Traits

Audiences

SQL Traits

Segment creates and merges user profiles based on a space's Identity Resolution configuration. Segment searches for identifiers such as `userID`, `anonymousID`, and `email` on incoming events and matches them to existing profiles or creates new profiles. These identifiers display in the Identities tab of a User Profile in the User Explorer:

JD Jane Doe jane.doe@gmail.com

Traits Audiences Events **Identities** Associated Accounts

Priority	Identifier	Value Limit	Blocked Values
1	user_id	1	Zeros and dashes, anonymous
2	email	5/Annually	Zeros and dashes, anonymous
3	group_id	5/Annually	Zeros and dashes, anonymous
4	ios.idfa	2/Weekly	Zeros and dashes, anonymous
5	android.idfa	2/Annually	Zeros and dashes, anonymous
6	android.id	2/Annually	Zeros and dashes, anonymous
7	ios.id	2/Annually	Zeros and dashes, anonymous
8	anonymous_id	5/Weekly	Zeros and dashes, anonymous

Let's say that two of your identifiers are phone number and email address. First, a user shows up on your website and registers with their email and phone number. The email is formatted in all lowercase, like `name@domain.com`, and the phone number is formatted as `+1-123-456-7890`. Then, the user calls your call center, and their phone number is formatted as `1 (123) 456-7890`. Sadly, the events from the call center will NOT get stamped onto the same profile which was generated from the website events, because the phone numbers don't exactly match. The same problem will occur if the call center agent types in the caller's email as `Name@Domain.com`. The mixed case email isn't an exact match with the lowercase email provided on the website, so the events from the call center won't appear on the same profile as the events captured on the website.

This approach to identity resolution is called “deterministic” because it requires exact matches (instead of “fuzzy” or “probabilistic” matches) on identifier values to unify events into a single profile. 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.

B2B Consideration: Group-Level Identifiers

If you're leveraging a CDP for B2B use cases, you'll want to make sure your identity resolution strategy also incorporates the grouping of user profiles into accounts or other entities that align with your business structure. Within Twilio Segment, you can use the Group Call⁴ to establish attributes (or **traits**) of group profiles, as well as many-to-many relationships between group and user profiles. You can also automatically associate user behavior with a specific group id by providing a new or existing group identifier inside the `context.groupId` element of any track call.

These relationships and behavioral associations enable the creation of account-level audiences⁵, which are especially useful for B2B audience building and targeting efforts. When building account-level audiences, you'll ultimately be able to surface a set of user profiles associated with both user-level and account-level criteria.

Compliant audiences

Unified profiles give you the ability to build highly-targeted Audiences based on criteria including events the user has (or hasn't) performed across touchpoints, computations on top of those events

⁴ footnotes.cdmbook.com/4

⁵ footnotes.cdmbook.com/5

(like `average_order_value`), and profile attributes like geographic region. But to stay compliant with privacy laws and policies, if you're sending an audience to an advertising or marketing platform, you must be sure to exclude from that audience anyone who says they don't want to be contacted for advertising or marketing purposes.

This exclusion does not happen automatically in Segment, since we focus on giving our customers the tools to enforce their interpretation of privacy laws and policies, rather than apply any interpretation on their behalf. This logic extends to our GDPR Deletion and Suppression endpoints (which enable you to remove and suppress further measurement of users with a `user_id` who have invoked their right to be forgotten), as well as our Privacy Portal⁶, which enables you to obfuscate or eliminate specific event fields with sensitive information from being sent out of Segment.

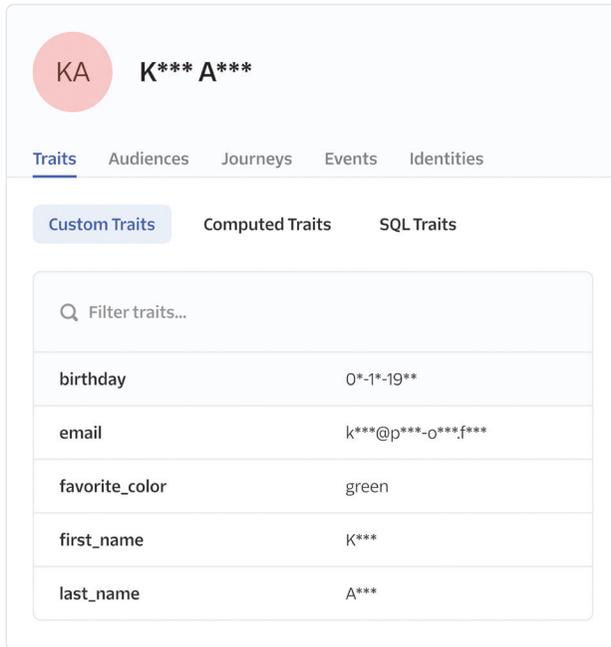
These tools collectively enable you to operationalize many aspects of your compliance requirements related to laws such as GDPR and CCPA, but the requirements themselves must be decided by you and your legal team. Once in place, those policies can help you craft rules for compliant audience-building, which you should be sure to share with anyone building audiences in Engage before they send those audiences to downstream tools.

Explicit preferences

Foundational profiles can be rich with useful information, even if you're collecting only a few additional profile attributes (called "traits" in Segment). And you don't need to do complex computations to turn that data into preferences to personalize user engagement.

⁶ footnotes.cdmbook.com/6

You can just explicitly ask your users for pieces of information, and then use what they say to tailor messages.



For example, let's say that as part of your app's registration process, you ask users for their birthday and favorite color. Those are explicit preferences you can store on their Engage profile, and then use to send them relevant messages like a "treat yourself for your upcoming birthday" email two weeks before their birthday, or an in-app message or modal showing the hottest products on your site that match the user's favorite color. The approach is surprisingly simple and also powerful. All that's required is thoughtfulness around the profile attributes you collect, and alignment between those profiles and creative approaches to engagement.

Customer Insight



Industry: Retail

Founded in 1946, Fender's worldwide 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.

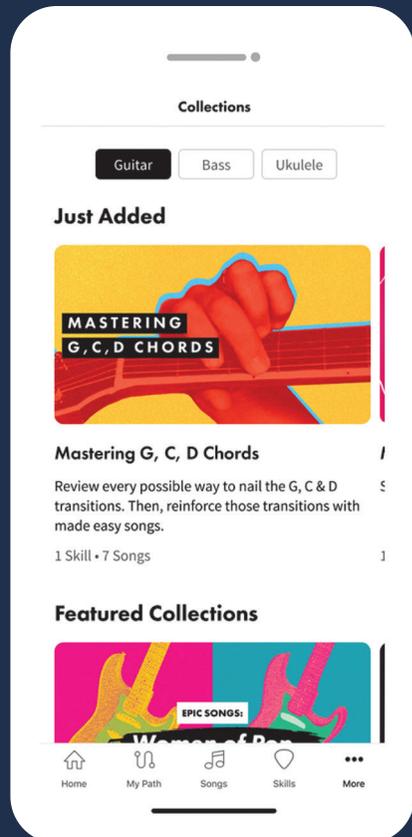
In 2017, Fender launched their subscription-based offering called Fender Play® – the complete learning app for guitar, bass, and ukulele.

The learn-as-you-go app is available as a monthly or annual subscription for iOS and Android users. With step-by-step guidance, students can study at their own pace while the app tracks their progress. Each bite-size session gradually increases in difficulty, upping a student's skill level while offering a rewarding experience.

To increase adoption and retention rates, Fender collects data using Twilio Segment to streamline and optimize the onboarding process. Using data collected via Segment, several customer friction points were identified. Leaning into this knowledge, they removed unnecessary steps in the sign-up process, driving a 5% increase in active paying users.

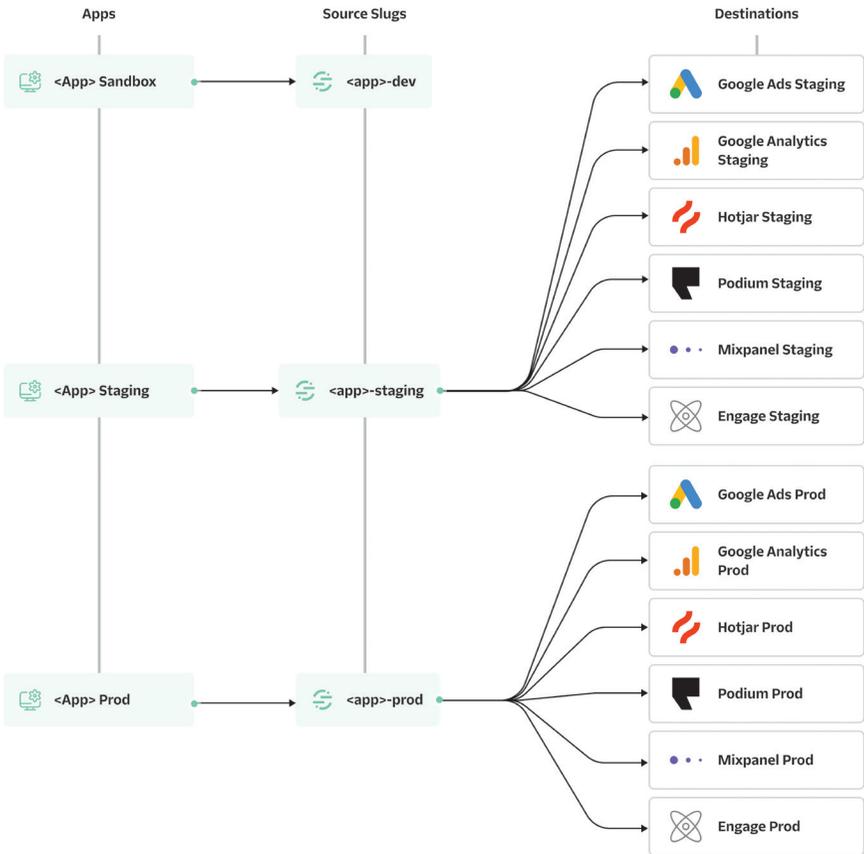
To further optimize onboarding they looked at their data collection screens and asked themselves if asking users about the instrument or type of music they wanted to play was necessary. To their surprise, most people didn't skip that step, so they kept those data fields in and cut screens users interacted with less.

By analyzing the data, Fender found that when users set their intention upfront, identifying their ability level and the instrument and genre they're most interested in, they were intrinsically setting themselves up for success. Users were choosing to engage in an activity they enjoyed (and would get personal satisfaction from) which made them more likely to continue practicing. Furthermore, providing gentle nudges when needed or kudos when a milestone is reached has contributed to reduced churn by 29%.



Thoughtful spaces

When getting started with Segment, we strongly encourage our customers to be thoughtful about how they’re setting up their tracking environments. Specifically, we recommend that you isolate your dev, staging, and prod environments to keep your prod data as “clean” as possible. Otherwise, you’ll need to “untangle” your dev and prod data in downstream tools such as analytics platforms.

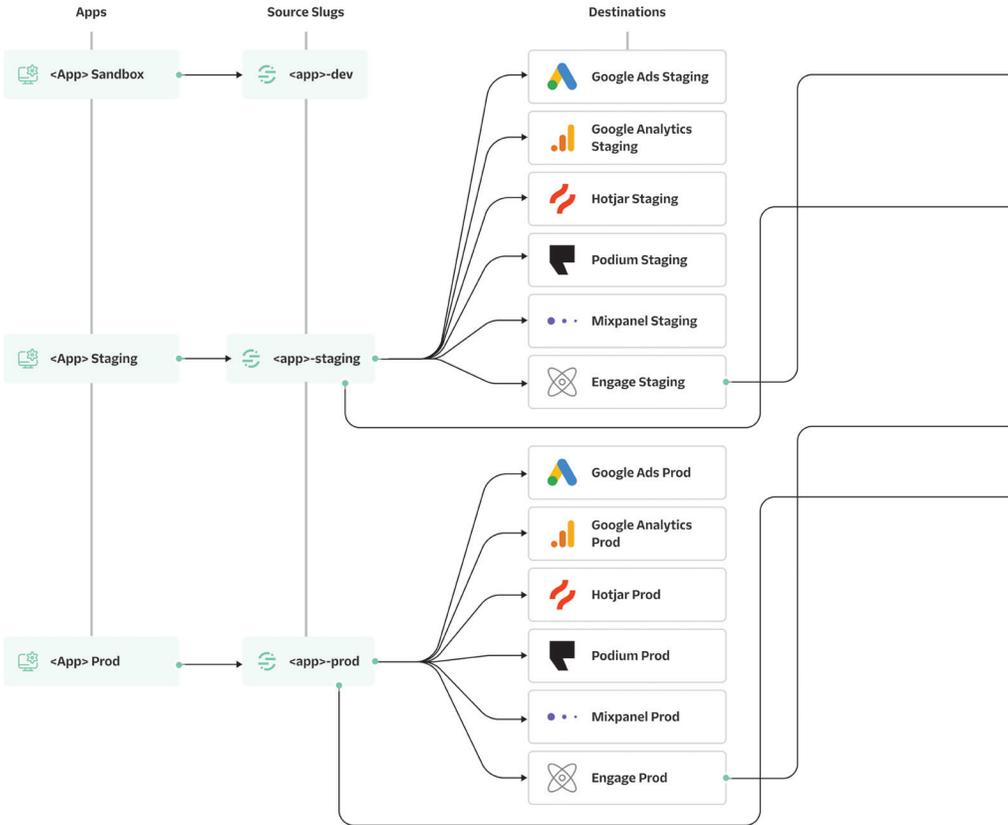


For example, if you simultaneously point a staging website and a live production website to a single Segment source, and then into a single downstream destination, the destination will contain all of the traffic from your test users and your actual users. This will throw off your calculation of key metrics like Daily Active Users, total orders completed, revenue generated, and retention. The prod data gets “polluted” with test and development data, and becomes untrustworthy to your stakeholders.

The same principle applies to Engage Spaces – which contain the identity resolution rules for a specific set of sources connected together to form unified profiles. If you combine your dev, staging, and production data together into a single space, you’ll mix actual user profiles with ones generated for development, testing, and QA purposes. This in turn will throw off your audience counts and may even cause good profiles to inadvertently merge with bad ones (once you merge a profile based on identifiers and identity rules you’ve established, you cannot undo that merge). That mixing and potential merging, in turn, causes distrust in the profile data, and reluctance from the marketing or product teams to activate those profiles with engagement.

For those reasons, we strongly encourage you to split your Spaces just like you’d split your sources. Doing so will give you well-defined boundaries between sometimes messy engineering and QA work and precious production data. If you’re looking for an easy way to classify your sources and spaces (and also manage access around that classification), check out Labels in Segment⁷.

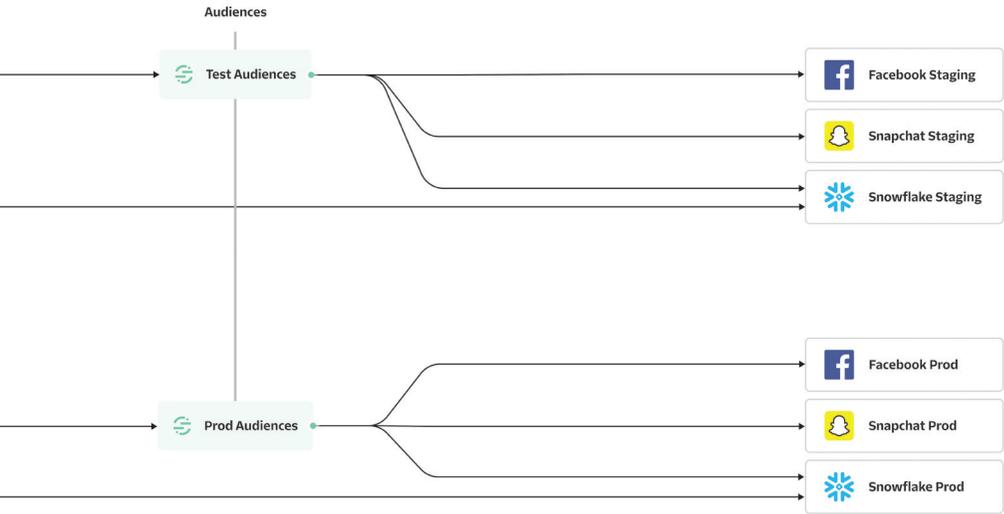
⁷ footnotes.cdmbook.com/7



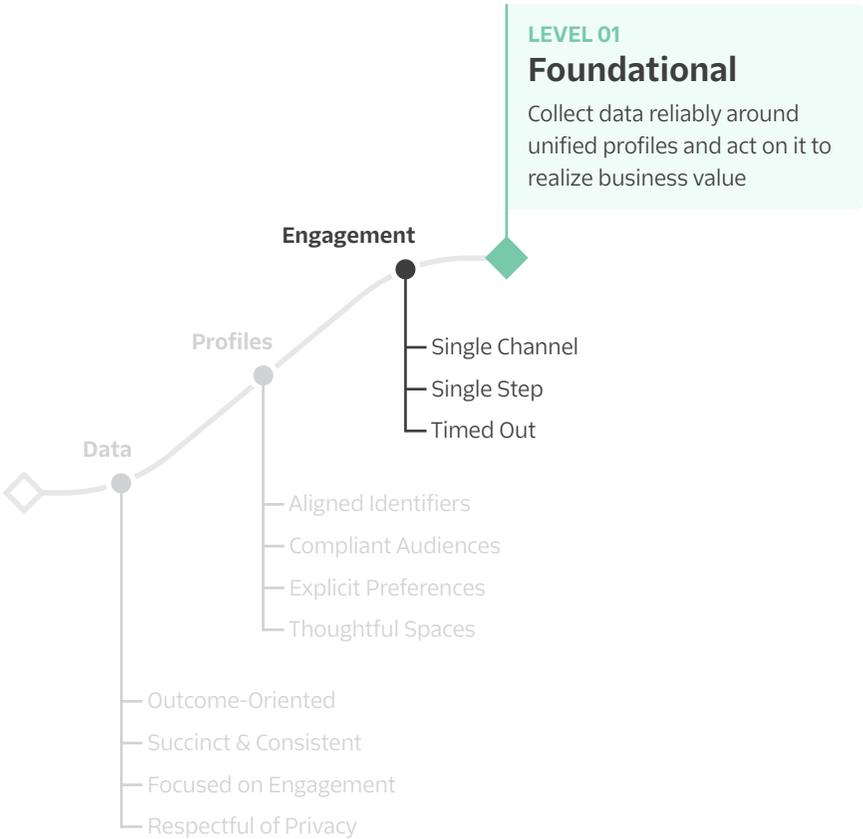
Additional Reading

For perspective on the common challenges and required capabilities of a successful first-party identity resolution strategy, check out our guide to Identity Resolution*.

* See QR code across from table of contents



Foundational Engagement



Building world-class customer engagement starts with choosing the right channels and making sure you're reaching the right audience. It may be tempting to jump right into implementing the ideal customer journey right away, but creating deeper, more impactful engagement with your customers requires a solid foundation.

Single channel

While omnichannel customer engagement can be particularly powerful for your business (more on that later!), when you're just starting out on your customer engagement journey we encourage you to *keep it simple*.

For example, let's say you wanted to message all customers who've added at least 5 pieces of clothing to their shopping cart in the past 30 days, but haven't yet completed a purchase. You might be tempted to build an audience of those specific cart-abandoners and then send that audience first to Facebook and Google Ads, then to an email tool, and finally to your push notifications tool – all with the goal of reducing cart abandonment and increasing conversions.

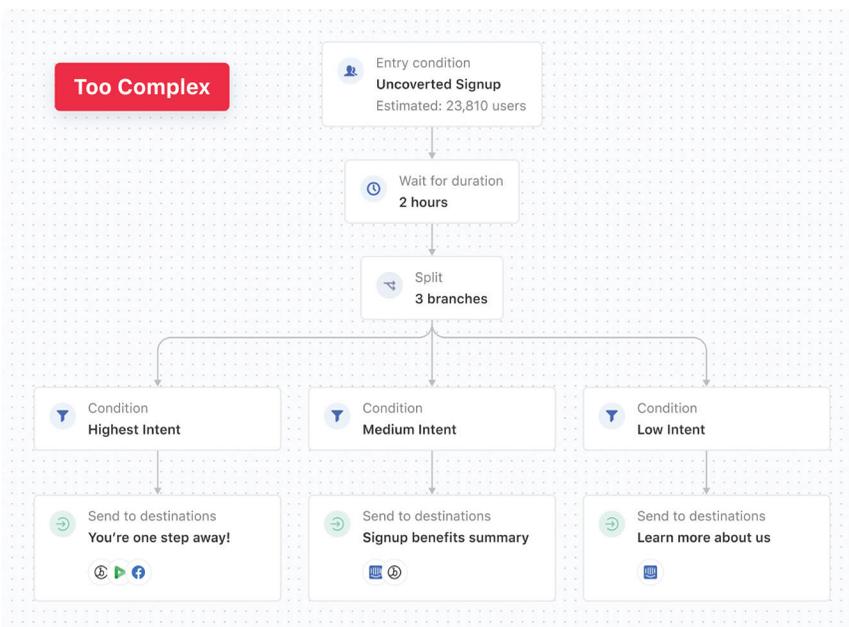
But why make it so complex? When you're trying to achieve foundational engagement, it's completely reasonable to just email all the customers in your audience who have an email address (and have consented to receive emails), and see how that moves the needle on conversions. Then wait a week, and do the same with an advertising campaign on Facebook. Then compare the lift between the two campaigns and see which one performed better.

For the sophisticated digital marketer, this might seem just too simple. Why not split the audience into A/B/C testing groups, and run each group through a different channel? You absolutely can, but in doing so, you're racing too fast towards advancing engagement before you've validated that single-channel campaigns work as expected. In working with thousands of customers, we've found that moving slowly and deliberately at first can surface unanticipated complexities with seemingly basic campaigns.

Then, once you've ironed out the wrinkles, you can scale with much more speed in the later stages of maturity.

Single step

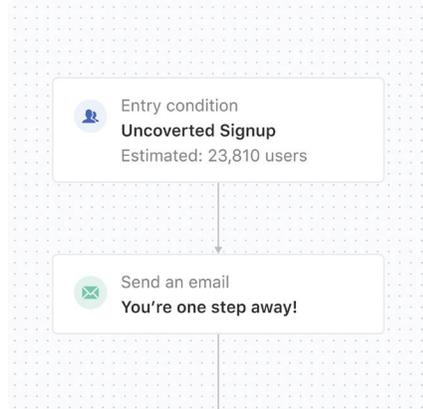
Just like it's tempting to send your audience to multiple channels for cross-channel activation, sophisticated CDPs make it equally tempting to build complex journeys with branching logic and nuanced conditions when activating that first audience. While establishing foundational engagement, do everything you can to *fight that temptation!* Journeys like the one below introduce unnecessary complexity when you're trying to achieve foundational engagement:



A multi-step journey; this is too complex for foundational engagement!

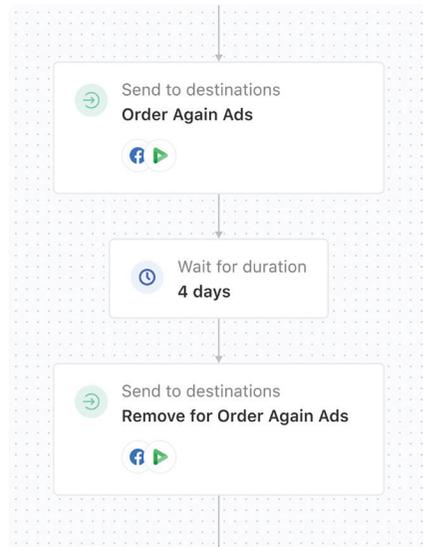
Just like you want to keep your channels simple during foundational engagement, you want to keep the steps in your journeys simple too. Single-step engagement is just that: the pairing of an entry condition and an activation action to move someone out of that entry condition. And

although single-step engagement doesn't leverage the full power of complex journey-building, it achieves something so much more important: a quick *measurable* win within your CDP!



Timed out

Audience suppression is essential to avoiding situations where you're wasting time and money targeting users who don't fit your target audience and are unlikely to convert. Oftentimes, audience suppression is applied to exclude customers who have *already* converted from campaigns. However, it is prudent to also "time out"



audiences as well – to exclude and suppress users based on how much time they spend in a targeted audience.

The longer a user spends in an audience, the more likely they will experience diminishing returns as you continue to target them. Suppressing users based on how much time they linger is a great way to curate your lists and amplify the success of each campaign, as you will always be targeting the freshest leads.

Assessing Foundational Success

Congratulations! You've made it through the first level of the customer data maturity model framework. By now, you should be familiar with the core concepts of customer data collection, unifying profiles, and building simple audiences which can be activated in real time.

We've covered a lot of ground, so let's quickly recap what you've learned:

- **Foundational Data** should be outcome-oriented (anchored on business goals), succinct, consistently named, focused on engagement, and aligned with your privacy policies.
- **Foundational Profiles** should have identifiers well-aligned across data sources, enable the creation of compliant audiences consistent with your privacy policies and consent preferences, contain explicit user-provided preferences to fuel engagement, and should reside in thoughtful spaces which reflect your dev, staging, and prod environments.
- **Foundational Engagement** focuses on simplicity by targeting only one channel at a time, single-step journeys, and timed-out suppression for advertising destinations so you can improve return on ad spend.

Next up, we'll move on to the Advancing stage, where we'll take our customer journey to the next level.

Case Study

Toggle

Industry: Financial Services & Insurance

Description: Toggle[®], a Farmers Insurance company, offers straightforward and customizable insurance products meant for the needs of today's modern consumer. To support the growing renters' marketplace, Toggle strives to offer best-in-class renters insurance, car insurance, and even Side Hustle[®] insurance and Identity Protection insurance.

How Toggle[®] sped up their customer data maturity using Twilio Segment

In a highly regulated and historically conservative industry like insurance, many businesses are lagging well behind in their ability to deliver data-driven customer engagement. Toggle was intent on bucking that trend, and adopted Twilio Segment to become the solid, scalable customer data infrastructure at the heart of their martech stack.

From the get-go, the Toggle team had data flowing into Snowflake, its newly adopted data warehouse, to gain actionable customer insights. With Connections, the team easily connected all of their

* See QR code across from table of contents

data sources and data destinations to collect customer data across every touchpoint, power their tools with the same customer data, and create a closed loop system for data flow. This includes mobile, website, their claims tool, payment tool, and customer support tool, as well as their marketing automation tool, analytics tool, and more.

Toggle then quickly turned on Segment Protocols data governance to align the entire company around a standard data spec. The team can now integrate new tools in minutes and have all customer data connected and cleaned with a few clicks of a button – all without relying on engineering resources.

With rock-solid foundations in place, they turned to Segment Personas (now part of Twilio Engage) to create in-depth customer profiles and simple personalized messages and campaigns.

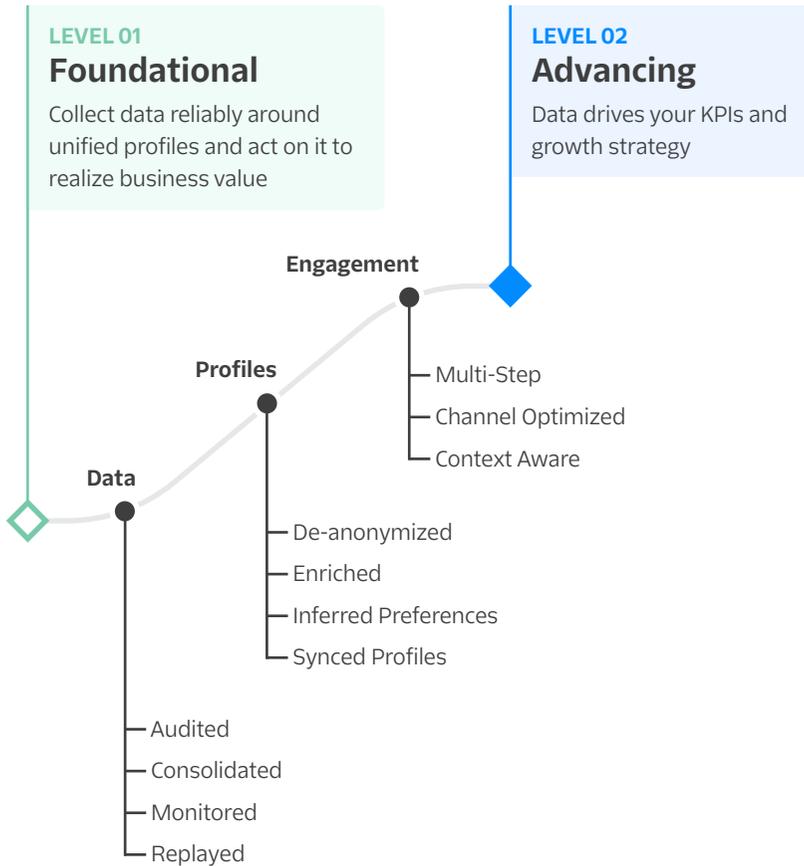
For example, to acquire new customers, the team sends hyper-targeted audiences built with first-party customer knowledge from Segment to model lookalike groups in ad platforms like Facebook and Google to retarget users with tailored ads based on actions prospects have taken on the site.

Similarly, once a user purchases a specific insurance product, the team can leverage customer data to send tailored content, curated newsletters, and complementary insurance offers. The team also uses Segment as a primary component for opt-in/opt-out management.

Toggle also uses Segment to provide better, faster, and more personalized customer support. By integrating claims data through Segment, the team is able to reduce friction and build a robust claims engine. For instance, after a claim has been filed, the team can send tailored 1:1 messages with updates on the claim and helpful links to help settle the claim faster.

Even though they're just getting started in their journey with Twilio Segment, Toggle has achieved remarkable business results in a short period of time: cost per acquisition has decreased by 67%, and they've increased Return on Ad Spend (ROAS) thanks to more targeted audience segmentation.

02 Advancing Maturity



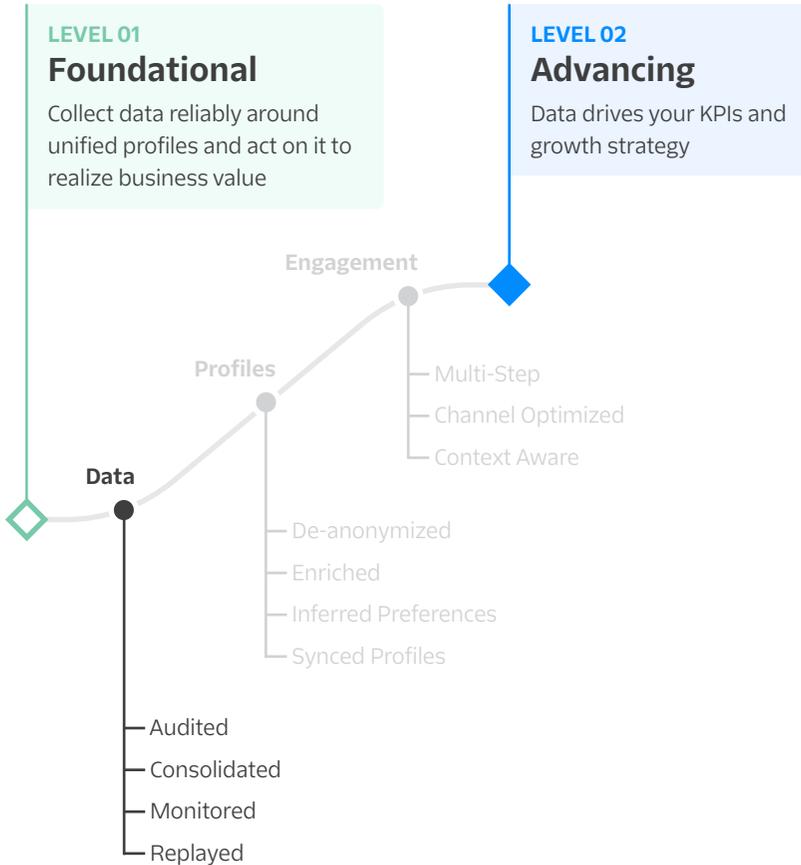
In the previous phase, we focused on laying the implementation foundations to begin engaging your customers on a basic level. Revisiting the house-building analogy, the next stage is where the house really begins to take shape and look like a proper dwelling.

The Advancing stage of data maturity is all about iterating on top of the foundation to continually assess and refine your data framework.

Where we previously sought to keep things simple while building the foundation, this stage is when you can start adding more complexity: whether by introducing more instrumentation, building more audiences, or joining profiles together in more creative ways.

The objective of this phase is to get to the point where you can more effectively activate your audiences so that your workspace becomes the primary driver for funnel conversion.

Advancing Data



As you advance along the CDP maturity curve, continued assessment and cultivation is the key to unlocking innovation and impact. While we revisit data at the advancing stage, the focus will be auditing what you've already built with an eye toward simplification and iterative improvement. This is not a one-time task but rather a continuous initiative—one that prevents the degradation of data over time, and helps foster a culture of continuous improvement and agile experimentation.

Audited

The more your CDP implementation grows, the more it runs the risk of becoming disorganized over time – no matter how meticulously planned your framework is or how strict your data governance practices are. To prevent this, you will need to proactively audit your data infrastructure on a regular basis. This audit is a three-part process: evaluation, diagramming, and schema reconciliation.

During evaluation, you'll ask all of your CDP stakeholders the following questions:

- What *works* in our CDP today and makes our jobs easier?
- What makes our jobs more difficult? What pains do we experience from our CDP?
- What CDP benefits do we want to realize in the next 6-12 months?
- What will happen if we don't achieve those benefits?

This is called the “Gains, Pains, Goals, Fears” exercise, and it helps align your cross-functional stakeholders around a shared understanding of the current state and future aspirations of your CDP. Use “design thinking” when asking these questions; put answers on sticky notes (or their digital equivalent) colored by team, and then identify the top 2-3 items per team. Ideally, you'll do this exercise with all stakeholders present, so everyone can hear cross-functional pain points and aspirations.

The next audit step is to whiteboard your current data architecture and identify any bottlenecks or inefficiencies. Are slow-interval ETL processes preventing fast time-to-action on your data? Are some systems

disconnected from others? Are manual steps required to complete a process which should be automated? Annotate the “current state” diagrams with its drawbacks, and then draw a “future state” diagram designed to alleviate those challenges. This future state should align with the goals you’ve identified in the evaluation exercise.

Finally, it’s time to dive into the details of your events and properties across sources – also known as your schema. When auditing your schema, you’ll want to take a “bottoms-up, then top-down” approach. Begin by looking at all of your events across all sources. How many uniquely-named events do you have, and how consistent are those events across sources? (Ideally, you’ll have 100 or fewer uniquely-named events fired consistently across similar source types.) Are there any exceedingly high-volume events? If so, how valuable are those events to the business?

Now that you’ve performed a bottoms-up analysis, it’s time to approach things from a top-down perspective. Without referencing any of your existing events, build a set of “golden funnels” which represent the idealized flow of succinct, consistent events measuring all business-critical user interactions. For example, you might have 5 funnels (Acquisition, Activation, Retention, Revenue, Referral), and 4-8 events within each funnel representing the major steps indicating funnel entry, passing the midpoint, and completion.

User Funnels & Actions

For each funnel, list the actions a user takes through the funnels

	Consumer	Merchant	Drivers		
	Initial Signup	Order Placement	Order Fulfillment	Subscription Product	B2B Signup
1	Application Installed	Screen Viewed	Order Received	Trial Started	Page Viewed
2	Application Opened	Product List Viewed	Order Accepted	Subscription Started	Form Step Completed
3	Signup Step Completed	Product List Filtered	Assigned	Subscription Renewed	Form Submitted
4	Signed Up	Product Viewed	Arrived	Subscription Cancelled	Merchant Account Created
5		Product Added	Food Ready for Pickup	Subscription Utilized	Store Creation Step Complete
6		Checkout Started	Pickup Completed		Store Created
7		Checkout Step Completed	En Route		Order Accepted
8		Promotion Displayed	Order Delivered		Order Ready
9		Coupon Added	[Vertical] Order Delivered	Convenience Grocery Alcohol Flowers Restaurant	Pass Enrolled
10		(First) Order Placed			Promo Created
	B2B Activation	Driver Signup	Driver Activation	Gift Cards	Notifications
1	Order Accepted	Page Viewed	Assigned	Item Added	Notification Sent
2	Order Ready	Form Step Completed	Order Accepted	Item Removed	Notification Delivered
3	Order Cancelled	Form Submitted	Order Declined	Item Updated	Notification Received
4	Pass Enrolled	Account Created	Arrived	Gift Card Purchased	
5	Promo Created	Account Reviewed	Pickup Completed	Gift Card Sent	
6		Account Approved	En Route	Gift Card Received	
7		Training Step Completed	Order Delivered		
8		Training Completed			

Existing Event	Related Funnel	consumer production	driver production	merchants production	All other production sources	% of total
1 m_card_view	?	7.6%	2.0%	1.1%		10.7%
2 m_card_view	?		10.4%		0.1%	10.5%
3 m_encounter_ab_test	?	0.1%	9.4%		0.1%	9.6%
4 m_network_request	?	6.3%			0.1%	6.4%
6 m_auto_order_active_polling_start				3.2%	0.1%	3.3%
9 m_network_response					0.2%	3.3%
10 fraud_rule_evaluation					0.1%	1.3%
12 m_order_status_card_view						1.2%
15 cx_pricing_order_quotes						1.8%
16 little_store_quoted_time						0.9%
17 m_identity_token_failure						0.9%
20 navigation_event						0.9%
21 m_scroll_menu_book						0.7%
26 proximity_request						0.75%
27 hidden_selection_discovery						0.5%
30 card_view						0.4%
31 m_chat_activate					0.1%	0.4%
32 m_card_click					0.2%	0.4%
33 m_search_item_view					0.1%	0.4%
34 m_cms_banner	?	0.3%			0.1%	0.4%
35 proximity_response	?				0.4%	0.4%

56%
of event volume
(30 of the top 50 events)
lacks alignment
with a funnel

With the golden funnels finalized, it's time to reconcile! For each uniquely-named event in your bottoms-up analysis, decide whether to keep it as-is, delete it, rename it to a golden funnel event, or split it into multiple golden funnel events. For example, if you have events called `Signup Button Clicked` and `Login Button Clicked`, you'll want to rename both to `Button Clicked` and set a property to indicate the `button_type` (signup or login). The Segment Professional Services team has performed this activity for hundreds of our largest customers, and in doing so we've found that the vast majority tend to either delete

or rename about 80% of their events. The end result is typically a 10x reduction in unique event names, more efficient architectures, and more satisfied analysts, marketers, and product owners.

Consolidated

Tracking more events than you need generates bloat, negatively impacting your bottom line in multiple ways. Analysts, marketers, and product managers waste time in downstream tools navigating complex schemas, your infrastructure incurs additional compute and storage costs for the data warehouse, and additional data engineers are often required to retrieve and clean messy data before it's considered operationally useful. For this reason, consolidating your events is of utmost importance to realizing healthy returns on investment in your CDP.

Your audit should have generated a blueprint for consolidation, and that blueprint should be reflected in your Tracking Plans, which at this point in the maturity journey can be upgraded to take advantage of Event and Property Libraries⁸. But the actual work of renaming events into a more efficient syntax can be time-consuming, and may require weeks or months of development time.

The good news is that there's a temporary fix you can employ to quickly show other stakeholders the benefits of a consolidated schema: use a custom Function⁹ to programmatically consolidate many similarly named events into a single event. (Segment's Transformations capability¹⁰ enables event renaming, but doesn't enable mapping of the inbound event name to a new event property, so Functions are required for a

⁸ footnotes.cdmbook.com/8

⁹ footnotes.cdmbook.com/9

¹⁰ footnotes.cdmbook.com/10

consolidation transformation.) Using this Function approach enables you to transform all new inbound events into your golden schema, and also enables you to replay older events to show the new schema for historical data in downstream tools – including Engage (more on Replays later in this chapter).

A word of caution: **don't fall into the trap of relying permanently on Functions to automate your schema cleanup!** This approach will certainly help in the short term, but it's only a stopgap, and it promotes the accumulation of additional technical debt upstream of your CDP, in the form of poorly named events instrumented into your code. To achieve true success at this stage of data maturity, you will need to address the underlying issues that caused event bloat in the first place. Work is certainly required, but the benefit is a fundamental shift in the way your team measures data – which results in powerful efficiencies when it comes time to measure the next set of outcome-oriented events.

Monitored

We've already spent the past couple of sections railing against a “set it and forget it” approach to your CDP, and now is the opportune time to talk about the importance of listening to your workspace. Monitoring your audit trail¹¹ and event delivery thresholds¹² helps you understand what's happening in your workspace at all times. Awareness of what is happening in your workspace gives you the ability to stay on top of changes and guard your workspace against disrepair.

¹¹ footnotes.cdmbook.com/11

¹² footnotes.cdmbook.com/12

To accomplish this, Segment offers comprehensive activity notifications – such as when events are blocked, a source or destination is created, a Tracking Plan is modified, and so on. “Listening” to your workspace means leveraging a feature called the Audit Trail, which captures everything that’s occurred in the workspace and displays the event stream. In Workspace Settings, make use of Audit Forwarding to forward this stream to one of your sources and connect it to destinations like a data warehouse – making it even easier to digest and monitor even the busiest of instances that generate thousands of events.

Audit Trail	
EVENT	TIMESTAMP
Destination Deleted	10/21/22 05:23 PM
Destination Disabled	10/21/22 05:23 PM
Computed Trait Run Failed	10/21/22 04:25 PM
Violations Detected	10/21/22 12:11 PM
Audience Modified	10/21/22 11:54 AM
Audience Modified	10/21/22 10:08 AM
Source Function Modified	10/21/22 09:16 AM
Source Function Modified	10/21/22 09:14 AM
Tracking Plan Operations Updated	10/21/22 07:26 AM



New Unplanned Event Blocked

A new unplanned event Blocked Event Test was blocked from `emilio_server` Source in the workspace `segment-success` and will not be sent to your Destinations due to your current Schema Settings.

New Unplanned Event Allowed

A new unplanned event Blocked Event Test was blocked from `emilio_server` Source in the workspace `segment-success` and will not be sent to your Destinations due to your current Schema Settings.

New Violations Detected

Below are new violations we've seen in your `emilio_server` Source since January 28, 2021, 12:00 am UTC.

Details

We've detected 1 unique violations in your `emilio_server` Source:

Event Name	Field	Type	Count	Last Seen
Product Viewed	<code>properties.user_id</code>	Required	1	8:27pm UTC

If you would like to allow these properties, you can add them to the [connected Tracking Plan](#).



⚠️ New Unplanned Event Blocked ⚠️

A new unplanned event called `Test` was blocked from the `emilio_server` Source in the workspace `segment-success` and will not be sent to your Destinations due to your current Schema Settings. You can add this event to connected Tracking Plan. You can manage your workspace notification settings [here](#). (edited)

✅ New Unplanned Event Allowed ✅

A new unplanned event called `Test` was blocked from the `emilio_server` Source in the workspace `segment-success` and will not be sent to your Destinations due to your current Schema Settings. You can add this event to connected Tracking Plan. You can manage your workspace notification settings [here](#). (edited)

⚠️ New Violations Summary ⚠️

Below are new violations we've seen in your `emilio_server` Source since January 28, 2021, 12:00 am UTC.

Event Name	Field	Type	Count	Last Seen
Product Viewed	<code>properties.user_id</code>	Required	1	8:27pm UTC

If you would like to allow these properties, you can add them to the [connected Tracking Plan](#). (edited)

Two examples of notifications generated atop Segment's Audit Trail

Replayed

As your CDP matures and data flows at higher volume and variety, you'll inevitably connect more destinations to unlock use cases like A/B testing or self-service analytics. When doing so, it's helpful to keep in mind that you don't need to "start from zero" in those tools, letting data accumulate only after you've turned on the tool to start receiving events. Instead, you can backfill historical data using a feature in Segment called Replays¹³.

Replays v. Historical Backfill

Replays require that the historical data already lives in Segment and isn't yet connected to a specific downstream destination. If your historical data needs to be backfilled into a Segment source and currently exists as a table in your data warehouse, you can make use of Segment's Reverse ETL capability¹⁴. With Reverse ETL, historical backfill is as easy as writing a SQL query and mapping the results table to common Segment methods in a Segment destination.

If your data isn't accessible in a Segment-compatible data warehouse, you'll need to write a script¹⁵ which traverses your historical data set and transforms each row into corresponding Segment API calls so you can ingest that data into Segment. If your historical data is in CSV format, you can make use of

¹³ footnotes.cdmbook.com/13

¹⁴ footnotes.cdmbook.com/14

¹⁵ footnotes.cdmbook.com/15

Segment's open-source CSV uploader tool¹⁶, or make use of Segment's Amazon S3 Source¹⁷, which provides sample code for a Lambda function you'll run on top of an S3 bucket. Regardless of the method, historical backfill is a critical component of data ingestion, and once performed successfully can make your Replays even more powerful.

Replays offer you a number of ways to advance your CDP maturity. Firstly, they enable you to foster a culture of experimentation at your organization without breaking the bank. Let's say your company wants to invest in self-service analytics. You can sign up for free trials with a few of the analytics platforms supported by Segment¹⁸, replay a few weeks of data (or a few months of just a few specific event names) into each platform so you don't exceed free trial limits, and then try out the UI with your own data to get a highly relevant side-by-side comparison. Then, when you're ready to go with one of the analytics platforms at scale, you can replay all of your historical data into the tool. And because you've already tested the tool with your own data, you can use it more efficiently with the full firehose of source data now streamed into it – speeding your time to value.

Replays also enable you to hydrate any number of Engage spaces with historical data (so you can build audiences, journeys, and computed traits on lookback windows), and they allow you to apply an event consolidation function to historical data before it enters an Engage space.

¹⁶ footnotes.cdmbook.com/16

¹⁷ footnotes.cdmbook.com/17

¹⁸ footnotes.cdmbook.com/18

Imagine you have a bunch of “messy” data in sources that creates challenges for audience-builders in Engage. If you replay that messy data through a consolidation function to a *new* Engage space, you can quickly show your stakeholders how a refined and consolidated schema would improve audience-building and computed traits. Their enthusiasm around the pristine data can in turn help rationalize an investment in engineering time required to normalize events at the source, so the consolidation function is merely a temporary stopgap.

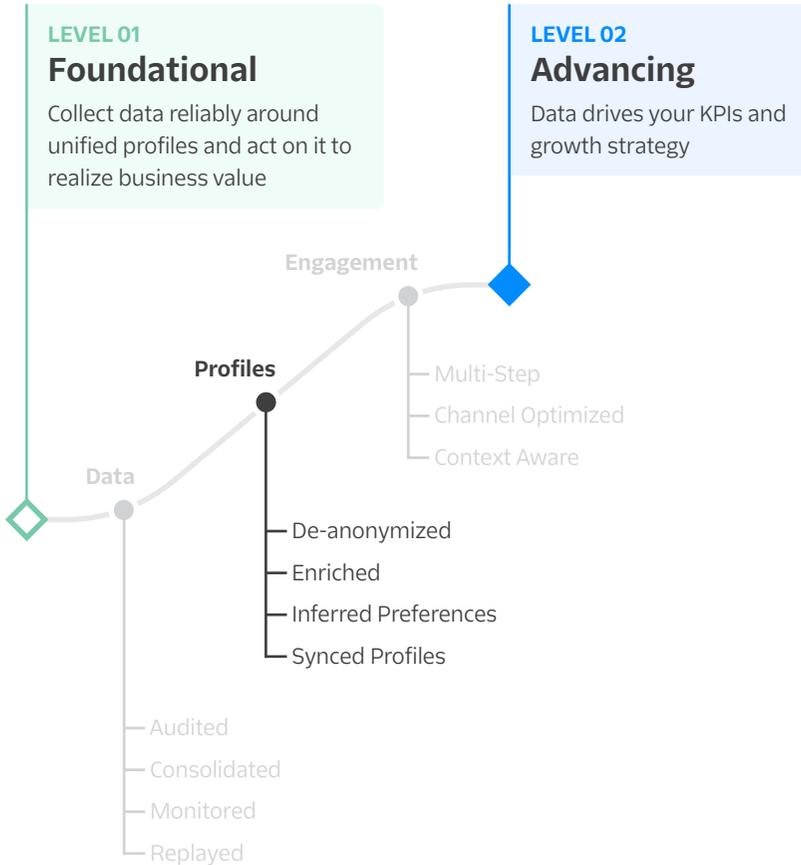
When performing replays, it’s helpful to keep an order of operations trick in mind. *First*, turn on the streaming data to the destination, and note the timestamp for the first event that appears in the destination. *Then* activate your replay, with the start time being anything you’d like, and the end time equal to the timestamp of that first event which appeared in the destination. This approach will ensure that you don’t duplicate events in your downstream destination.

B2B Consideration: **Group Call Support in Destinations**

If you're making use of Group calls in Segment, make sure you first check the Destination methods comparison chart¹⁹ before sending Group calls to your desired downstream destination. If Group calls aren't supported in your destination, you can often construct workarounds using Segment functions (to remap the group data back to user profiles in the downstream destination), or by rethinking your use case for account-level data in the downstream destination. In both instances, Segment's Professional Services team can help devise and build solutions that enable realization of your B2B use case. These solutions sometimes include reducing the amount of raw account-level data sent downstream, and instead making use of account-level audiences to send lists of associated user profiles.

¹⁹ footnotes.cdmbook.com/19

Advancing Profiles



It's time to talk about refining Profiles, which involves merging them effectively, enriching them with additional traits, and even gaining deeper insights into your customers with inferred preferences. Adding increasing levels of complexity to your instrumentation expands the types of Audiences you can build, allowing you to think more creatively on how to activate different segments.

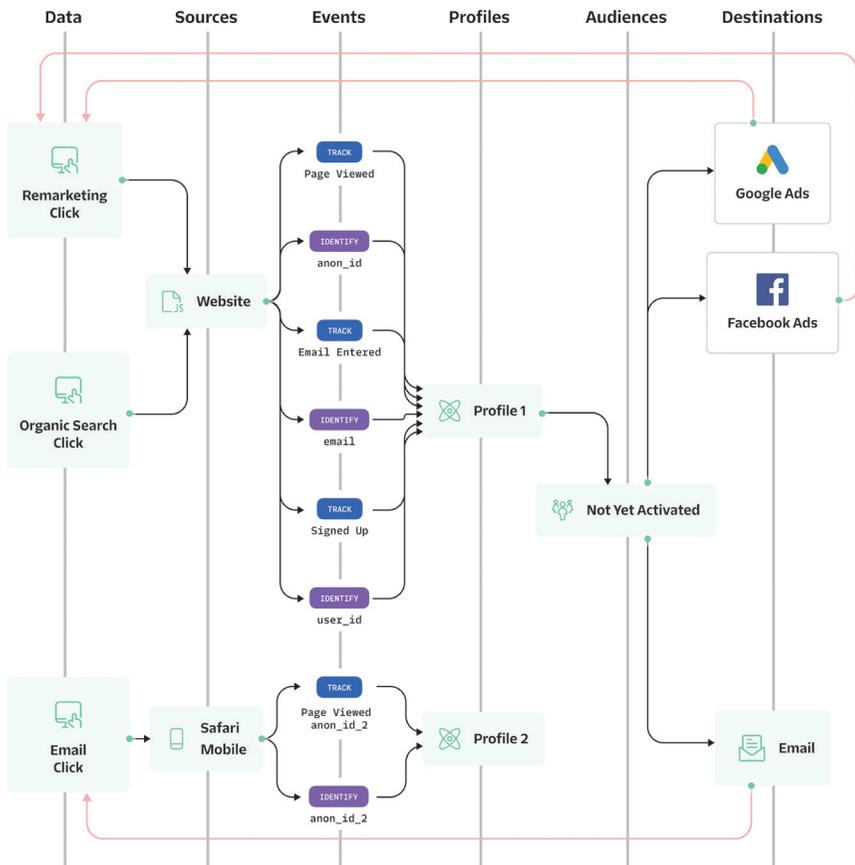
De-anonymized

Anonymous visitors account for upwards of 98% of all website visitors. Maybe they're visitors who haven't converted yet. Or perhaps they have converted, but subsequently cleared their cookies after logging out. With a CDP like Segment, you can reconcile anonymous visitor data²⁰ with your known visitor data to unify profiles and avoid gaps as you analyze customer behavior.

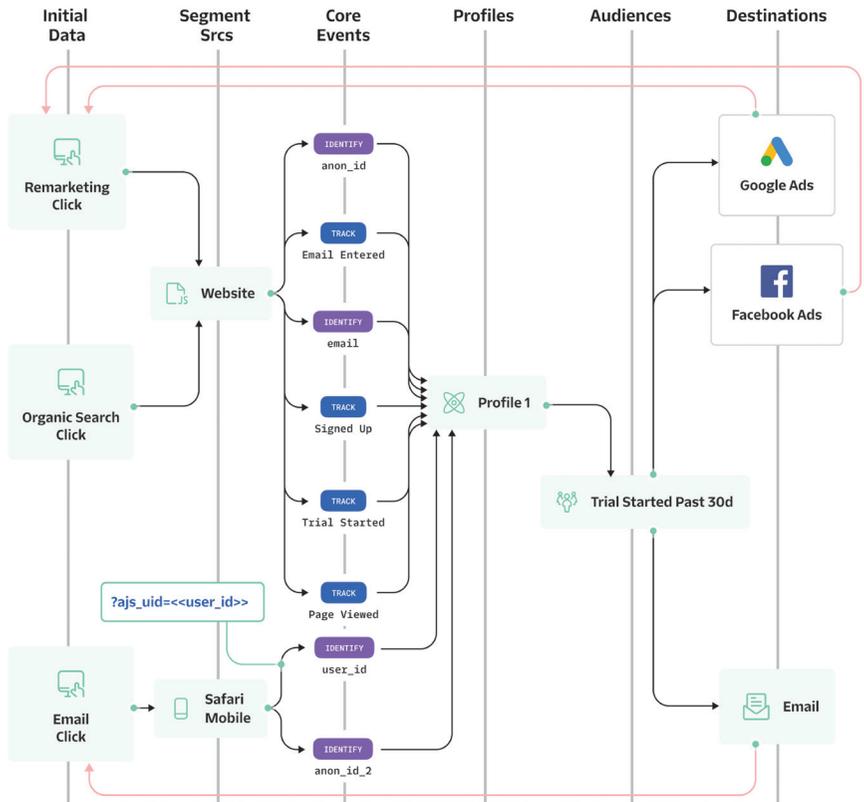
However, even after you successfully identify a user and associate their previously anonymous activity with their identified profile, you still run the risk of that user showing up as a separate profile when they visit from another device or browser and haven't yet signed in again. This is most common with emails clicked from mobile devices, when initial registrations took place on a laptop.

²⁰ footnotes.cdmbook.com/20

In the example below, a user first visits a website via organic search, then signs up, and then revisits the site via remarketing click. At this point, a single profile captures two visits. However, when the user clicks on an email via their mobile device and doesn't yet sign in, their browsing behavior on the mobile device is tied to a completely separate profile:



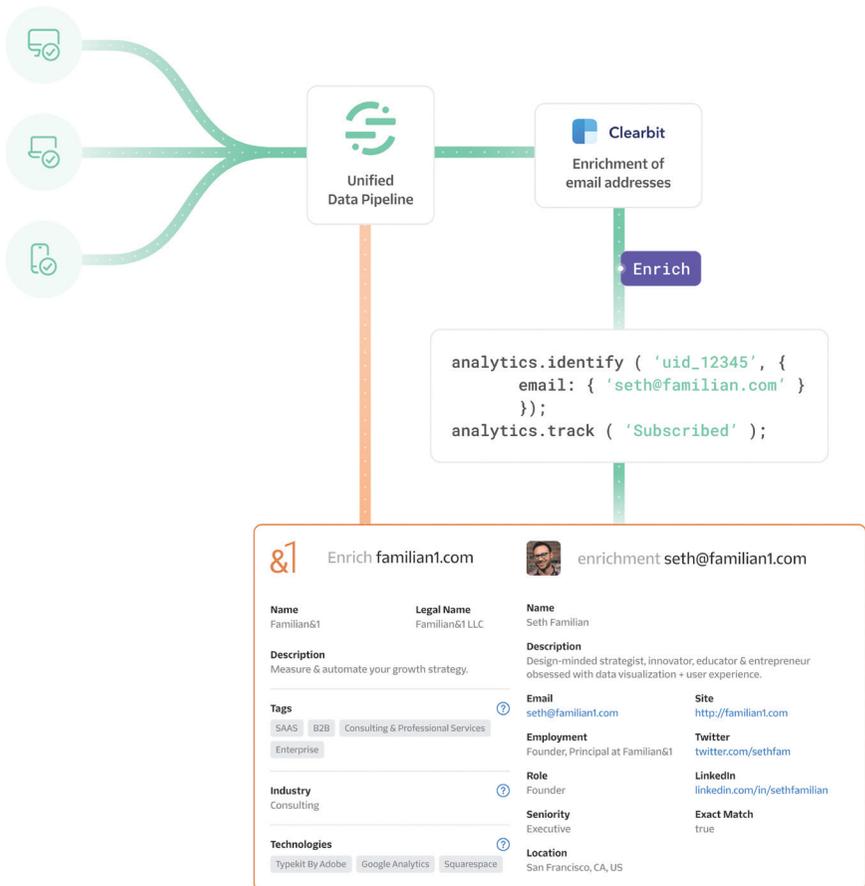
Segment customers solve this common challenge by taking advantage of Segment's Querystring API²¹, which enables the inclusion of a simple parameter at the end of a url that prompts Segment to fire an identify or track call when that url is loaded on a page containing Segment's analytics.js library. In the example below, the `ajs_uid` querystring value tells the page loading the url to fire an identify call with the `anonymous_id` on the page aliased to the `user_id` provided in the querystring. This unifies the otherwise split profiles and de-anonymizes the user's behavior immediately.



²¹ footnotes.cdmbook.com/21

Enriched

If your leads come from forms or account signups, you want to know who's knocking at your door. While Segment's identify call can give you powerful demographic information like the referral source, campaign information, and device type, enriching your core Segment data using a third-party vendor like Clearbit can push you into warp drive.

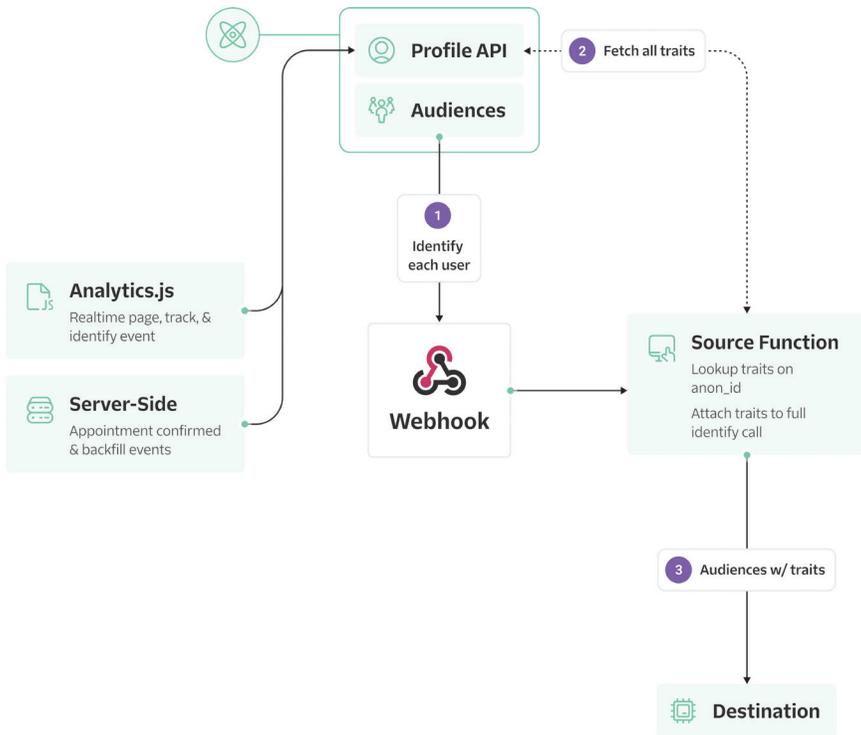


Using Segment in tandem with a third-party enrichment tool can automatically turn a work email address (or corporate domain name) into a full customer profile with company, job, and contact information, which unlocks a whole different dimension of personalization. For example, by knowing where each company is headquartered, you can ensure your messaging is sent in the appropriate language. Another strategy you can employ is to leverage the Clearbit “Seniority” field to ensure your messages are targeted to the appropriate stakeholder level.

Synced profiles

Segment is an extremely powerful CDP that powers the customer data infrastructure for large enterprises, which means that an active implementation may be ingesting events and traits from tens of millions of tracked users per month. The problem is that your downstream tools may be much more limited in how much they can ingest; perhaps instead of millions, your email platform may be limited to ingesting 500,000 new users per month.

Larger Segment customers solve this challenge by choosing not to send all of the raw traits and events directly to destinations with tight profile limits. Instead, these customers rely on Segment to ingest all of the raw data, resolve profiles, and build audiences. Then, those customers choose to send only specific audiences downstream, and they take advantage of a custom-built Source Function to “sync” the full profile downstream only when a new Audience member is identified by Engage.



The Source Function above looks up an Engage-identified user on the Segment Profile API, and then enriches the identity call with any number of profile traits (like `first_name` and `favorite_color`) for more personalized campaign-building downstream. “Syncing” the profiles in your downstream destinations with this Function is a far more efficient way to identify only the users who you’ll be targeting in the destination, and helps avoid overages generated from the raw source data connected directly to the destination.

Some of Segment’s most advanced “Actions Destinations” support this capability (called “Trait Enrichment”) directly in Segment’s UI. As more and more Segment destinations become Actions-based

(an evolution fueled by Segment’s strong network of partnerships), Trait Enrichment will be more widely available across the platform. In the interim, an enrichment source function like the one described on the previous page serves as a helpful stopgap.

Inferred preferences

What if you could automatically discover the channels that customers preferred, and engage with them accordingly? It’s actually not a far-fetched possibility. As shown in the diagram below, you can use Computed Traits²² in Segment Profiles to calculate which `utm_source` or `utm_medium` a user engaged with most frequently...



Most frequent

Determine the most frequent property value of an event

NAME	MOST FREQUE...	MOST FREQUE...
AF	A*** F***	Email
EL	E*** L***	Email
IH	I*** H***	Email
ZR	Z*** R***	Facebook
WT	W*** T***	Google
MS	M*** S***	Email
DW	D*** W***	Google

Select an event name

Page Viewed

Add Conditions

Select an event property

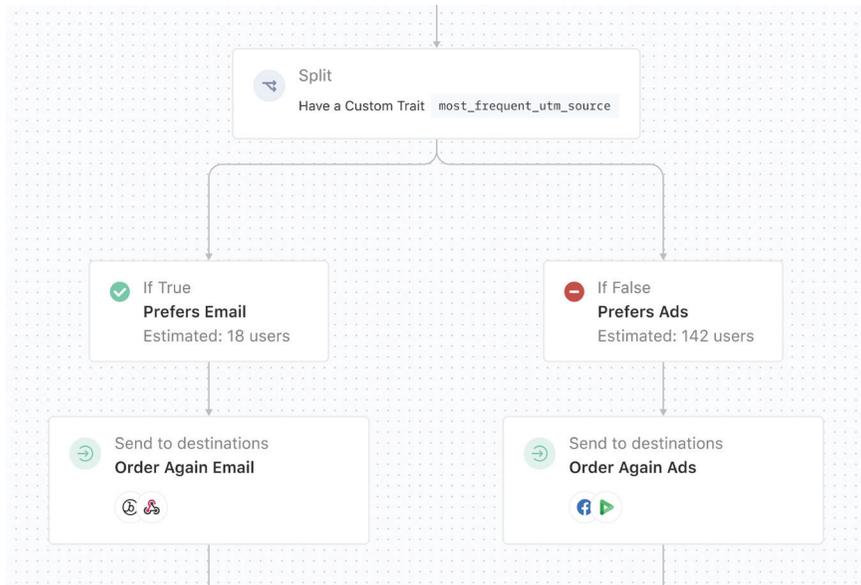
campaign.source Text

Minimum frequency 3 ⓘ

Include anonymous users

²² footnotes.cdmbook.com/22

...and then use that trait in a complex Journey to send users communication via the most relevant channel first:



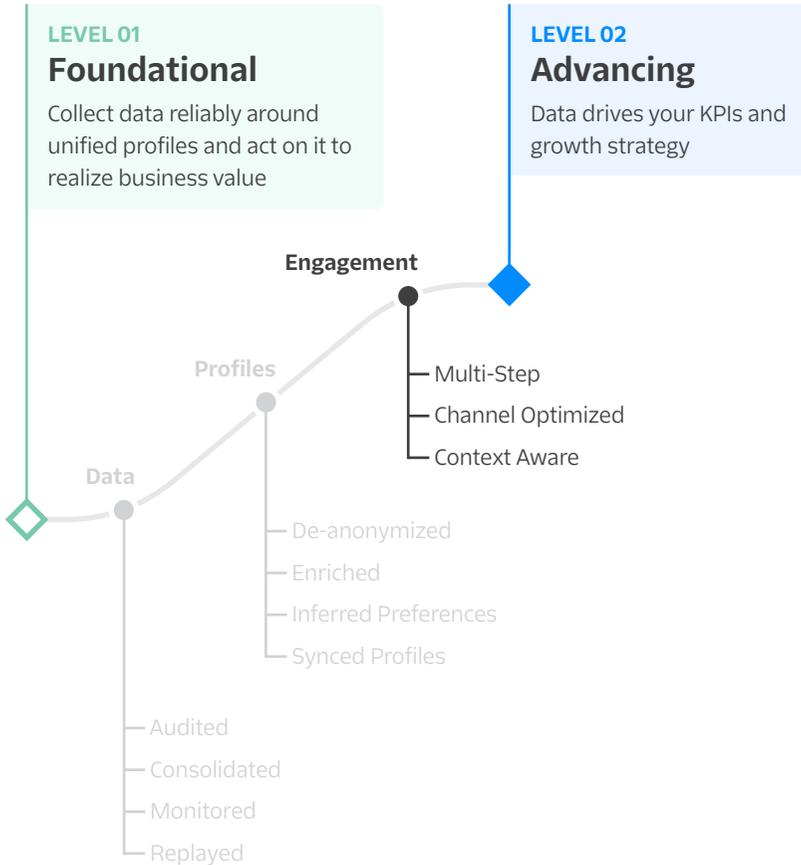
In the Journey shown above, we've split the users based on whether their Most Frequent UTM Source was email (true) or not email (false) and we then engage with those users accordingly. Doing so enables us to optimize our ROAS on those who we've inferred prefer advertising, and also enables us to optimize email open and click rates by sending emails only to those who indicate that's their preferred channel.

B2B Consideration: **Account-Level Computed Traits**

For B2B applications, Segment offers account-level computed traits. These computations run on top of events ingested into account profiles, and produce a net new Group call with the computation stamped onto the account's profile. These account-level computations unlock powerful B2B use cases, including insight into aggregated behavior across all users associated with a specific account, or surfacing the first or most recent user performing activity related to an account.

Some of Segment's most advanced B2B customers make use of account-level computed traits to trigger additional identify calls and other enrichments in specific user profiles related to the account – such as the “account representative” (for insurance policies) or the “lead passenger” (for airline ticketing).

Advancing Engagement

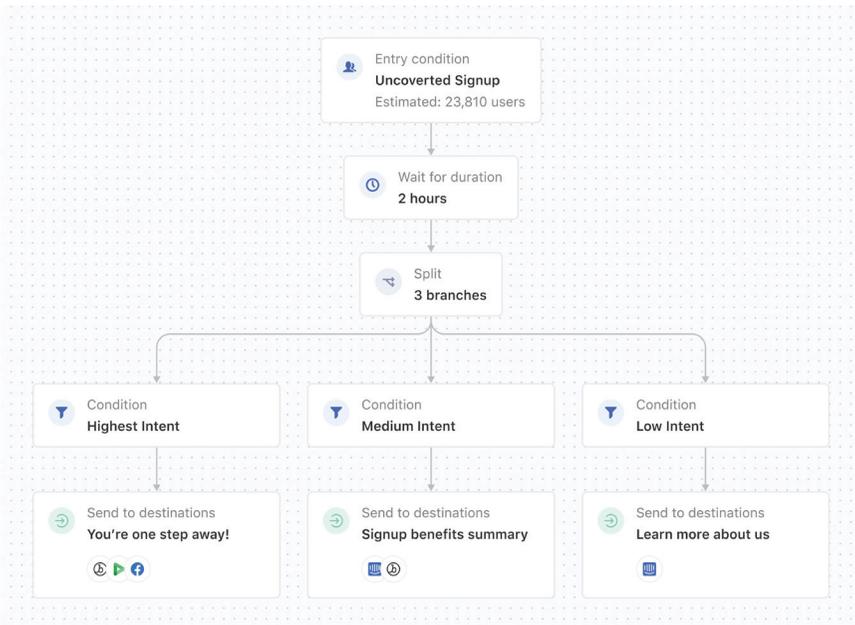


Advancing Engagement builds on top of the basic single channel, single step customer journeys laid out in the foundational phase. Here, we will be working on adding more complexity and personalization to reach your customers in more engaging, meaningful interactions.

Multi-step

The customer journey is rarely linear. As your customer engagement capabilities become more mature, you can take personalization to a whole new level with what we refer to as “Journey branching.” With Journey branching, marketers can customize paths in Journeys for users based on their unique attributes and actions.

An excellent example of Journey branching is the unconverted signup flow considered too complex for Foundational engagement, but fantastic for Advancing engagement, since it leverages inferred preferences across branching logic:



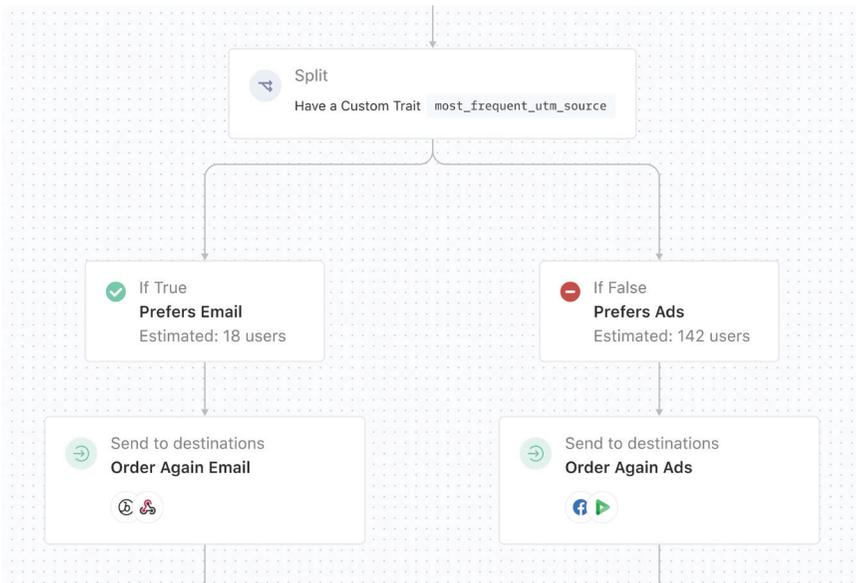
In this Journey, we've split the experience into three branches based on the inferred "intent" someone indicates in their most frequently viewed pages:

- Highest-intent users visited the "Get Started" page most frequently
- Medium-intent users visited the "Landing" or "Products" pages most frequently
- Low-intent users visited the "Home" or "About" pages most frequently

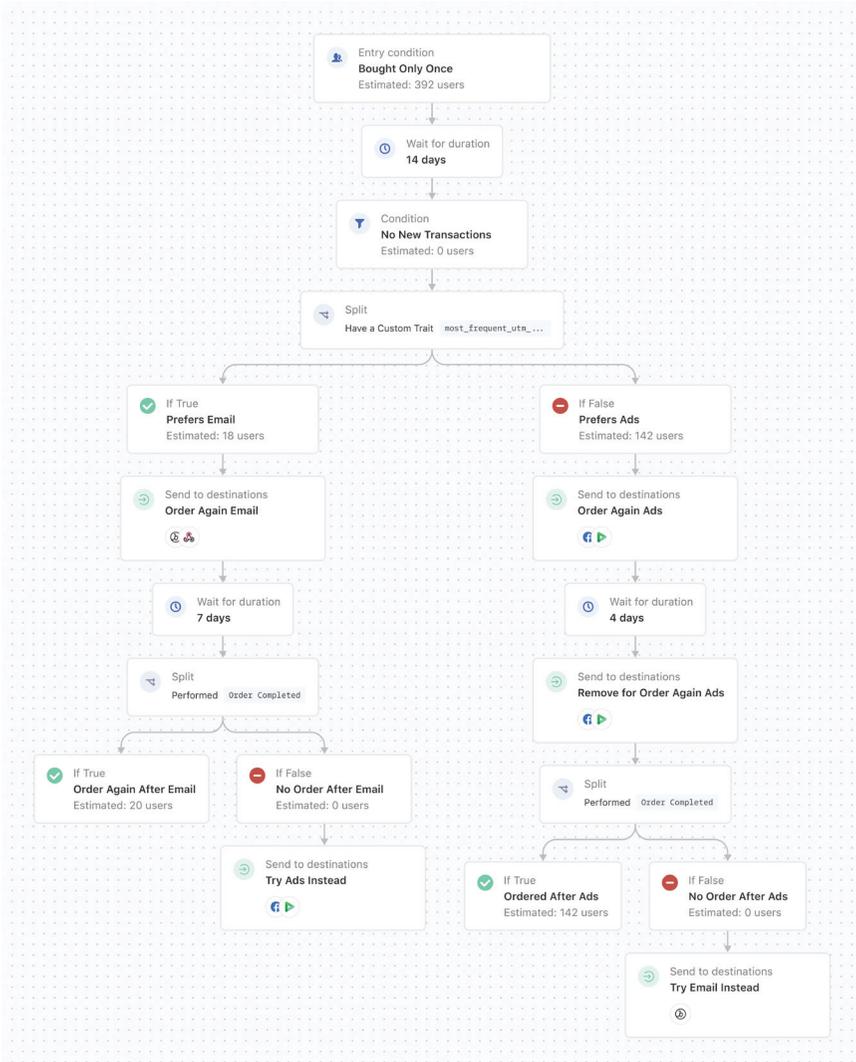
For each level of intent, we've engaged the user with a different call-to-action and with a different blend of channels. In doing so, we've generated hyper-relevant experiences that personalize the entire Journey for users, instead of delivering one-size-fits-all messaging.

Channel optimized

Optimizing your communications around the channels where a customer is most likely to engage is a powerful tool in your personalization toolkit. An excellent example of channel-optimized messaging starts with the inferred preference flow discussed at the end of the Advancing Profiles section:



This flow is actually just one part of a much larger multi-step journey which anchors first on the user's preferred channel (based on inference) and then falls back to the other channel if the first one doesn't drive a conversion.



While highly complex, this Journey also demonstrates the true power of Engage and its ability to optimize for messaging across channels. It also illustrates a number of the concepts we've discussed so far, including timed-out advertising, inferred preferences, and

multi-step Journeys. Consider it an inspiration for the sophisticated engagement you can achieve with Engage.

Context aware

In the scope of advanced customer engagement, you should also be thinking about how to use data to enhance the level of personalized service when you directly interact with customers.

To accomplish this, you will need to equip customer service agents with the contextual data they need to understand who your customers are, how they've historically engaged with your business, and what their needs are. With a customizable contact center like Twilio Flex, you can customize the agent dashboard to surface data from Segment's Profile API²³ to give your agents the full context into each customer's engagement across all channels. The diagram below shows an example from a fictional airline, Owl Air:

The screenshot displays a comprehensive customer profile for Jesse Cordoba. Key details include a Gold status, 250,780 miles, and a contact number of +18650813546. The journey timeline shows a sequence of interactions: flight searches for San Francisco to Singapore and San Francisco to London, followed by completed web sessions. The customer's contact information, including phone and email, is listed. A list of activities shows recent calls and chats, with notes on product requests and cancellations. The chat window on the left shows a recent conversation with the customer.

²³ footnotes.cdmbook.com/23

In the Owl Air example (p.82), Segment was able to capture what flights the customer was searching for from their web session so that when the customer asks for help with accommodations, the agent has the information to intuitively understand exactly where help is needed. Segment's Profile API merged the data from the CRM *and* the customer's web activity to give the complete view of the customer.

Oftentimes, customer service agents will only have one source of data available, such as a CRM or helpdesk, but this will only result in gaps where they may not have the information on hand to help the customer. To avoid frustrating the customer with lengthy back-and-forth conversations, you'll need to integrate every source of customer data to build a complete picture of the customer.

In a real-world application, you can add additional channels as data sources or surface Computed Traits from enriched Profiles to provide more advanced levels of engagement. With this level of context, you can actually personalize engagement on demand to deepen customer relationships.

Additional Reading

To learn how to integrate Segment into your Twilio Flex instance, check out our blog, [Personalize your Twilio IVR with Segment*](#).

* See QR code across from table of contents

B2B Consideration: **The Links Endpoint**

When integrating profile data into applications such as Flex, Segment's B2B customers often make use of the /links endpoint of the Profile API, which surfaces all accounts linked to a user, or all users linked to an account. Traversing all accounts associated with a particular user ... and all users associated with those accounts ... enables a salesperson or B2B marketer to see the complete picture (or "graph") of a particular person within an organization (how many different opportunities they're associated with, how many active accounts they touch, and how many others they potentially influence).

The same API can also give us insight into the traits or audience memberships associated with those linked entities. So when someone from a particular company interacts with us via Flex, we can see that the company itself is part of an account-level audience called "high-potential upsell this quarter."

Assessing Advancing Success

Nice work! You've just completed step two of the customer data maturity model framework. By now, you should be familiar with the advanced concepts of customer data collection, adding rich personalization to your customer engagement strategy, and sending multi-step, contextual messaging to your audiences. We've covered a lot of ground, so let's quickly recap what you've learned:

- **Advancing Data** is regularly audited & consolidated, monitored via your Activity Feed, and replayed into new tools (or Engage) to cultivate a culture of experimentation.
- **Advancing Profiles** are de-anonymized via Segment's querystring API, enriched using destinations like Clearbit, synced with downstream tools that have tight usage limits, and infer preferences using Computed Traits to drive more powerful engagement.
- **Advancing Engagement** targets users across multiple steps and channels while leveraging inferred preferences to deliver more complex and robust Journeys. That same engagement makes use of the Profile API to ensure your customer service interactions (whether in the call center or at retail) are context aware of the user's previous behaviors across all of your Segment touchpoints.

The next and final stage is the Adaptive level of maturity, where we'll pull everything together to show you how to become a data-driven category leader in your industry.

Case Study



How IBM advanced their organization's data strategy to drive new revenue

When IBM first adopted Twilio Segment at the start of their customer data journey, the team decided to start small and prove value along the way.

IBM scoped the challenge of enabling the product, growth, and sales teams for a single product line with a common data infrastructure that could give visibility into the customer journey and enable more personalized campaigns.

First, to achieve a unified view of the customer, the IBM Cloud team decided to create a best-in-class technology stack powered by Twilio Segment.

The team uses Twilio Segment for data collection, standardization, and activation due to its ability to scale across multiple product lines and provide consistent data across a variety of tools their teams wanted to use. IBM Cloud integrates analytics, A/B testing, enrichment, and data warehousing tools with Twilio Segment.

Downstream, IBM leverages the artificial intelligence of IBM Watson AI products, such as Watson Studio and Watson Natural Language Classifier, to identify deep customer insights. This process would have taken months of time and a significant amount of people investment without Twilio Segment.

For the first initiative, the team instrumented Twilio Segment on IBM Cloud to capture all user behavior associated with that product. Then, they piped the data into Amplitude for the product and growth teams to analyze user behavior and to Salesforce for the sales team to get more visibility into potential product upgrades.

Using the analysis, the product and growth team was able to identify where a lot of customers were becoming disengaged and where potential upsell opportunities existed. Narrowing in on these opportunities, they used data in Intercom to trigger personal notifications to better engage inactive customers and cross-sell new products.

“We had a vision of being able to track everything consistently,” said Peter Ikladios, Director of Growth and User Engagement at IBM. “We knew we had to start small. The starting point was delivering better insight into our customers and communicating with them in a seamless way. We set up a pilot program with our nurture tool and Twilio Segment to test the viability of this vision.”

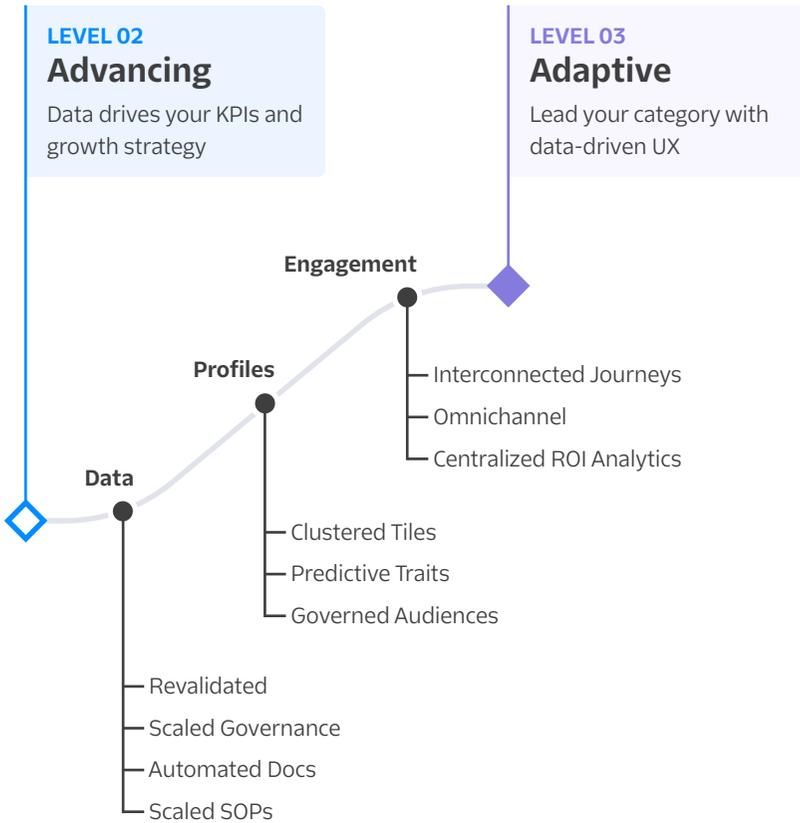
By using Twilio Segment to power its nurture streams with customer data, IBM Cloud was able to tailor its messaging to individual customers based on each customer’s history. The success of this pilot encouraged the IBM Cloud team to conduct a scaled roll-out.

“Twilio Segment allows us to handle the disparate data sets we were using. We can clearly understand what users are doing with our product. It doesn’t matter where you look at the data – with Twilio Segment as the data foundation, it’s always the same. We can clearly understand the data lineage and trust the information that’s being displayed across our business.”

The initial nurture pilot with Twilio Segment and Intercom led to a 70% increase in revenue over a three-month test period.



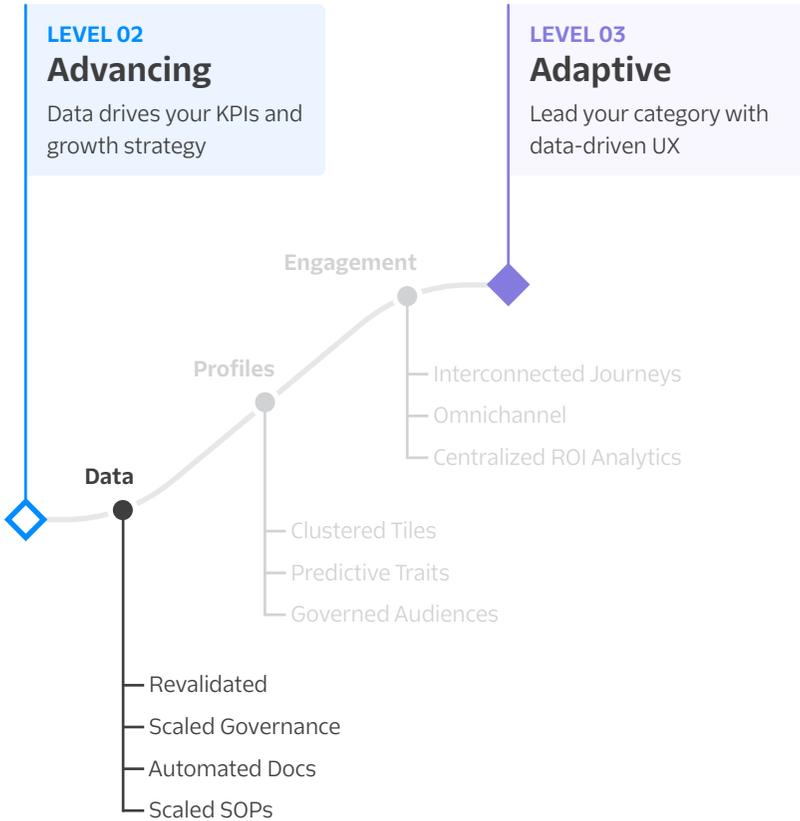
03 Adaptive Maturity



Let's return to our analogy and retrace the steps we took to build our "theoretical" house. In the previous stage, we built the rest of the building on top of solid foundations – the house starts *looking* like a house. However, even when it's structurally sound, a house doesn't really "wow" the neighborhood (and the real estate market as a whole) unless it's finished with beautiful and innovative details which set the standard for all houses around it.

The house analogy culminates in the Adaptive stage of data maturity: becoming a category leader by leveraging your CDP to create personalized customer experiences at scale. Creating personalization at scale occurs when you unify data across all touchpoints, deploy automation for data governance processes, and implement predictive analytics. Following the path to success in the Adaptive stage of data maturity puts your organization in a position to lead your industry sector with omnichannel engagement.

Adaptive Data



At the most advanced stage of data maturity, category leaders are expected to execute data governance processes at scale. Executing at scale requires both the standardization of data governance processes within a mature organization, as well as automating certain technical components of data governance when possible.

Revalidated

Large and mature companies have usually built highly complex CDP infrastructures with hundreds (and sometimes even thousands) of sources, destinations, custom functions, audiences, journeys, and active users. However, even the most innovative infrastructures can miss opportunities for efficiency and simplicity. That's why it's critical to put in place an *ongoing, repeatable audit process* for your entire CDP infrastructure. Otherwise, you'll generate unnecessary risk and cost.

For example, Segment's Professional Services team recently helped audit the workspace of an enterprise customer quite advanced in its data maturity, but still facing inefficiencies in its architecture. By making two different adjustments to its data architecture, **the team saved 1.6 billion API calls per month**, and eliminated 120 repeaters and follow-on sources. Another adjustment saved them an extra billion fewer rows in their data warehouse.

No matter how mature your organization, take a step back to examine the architectural decisions you've made in the past, and revalidate whether there's a better way to do it.

Scaled governance

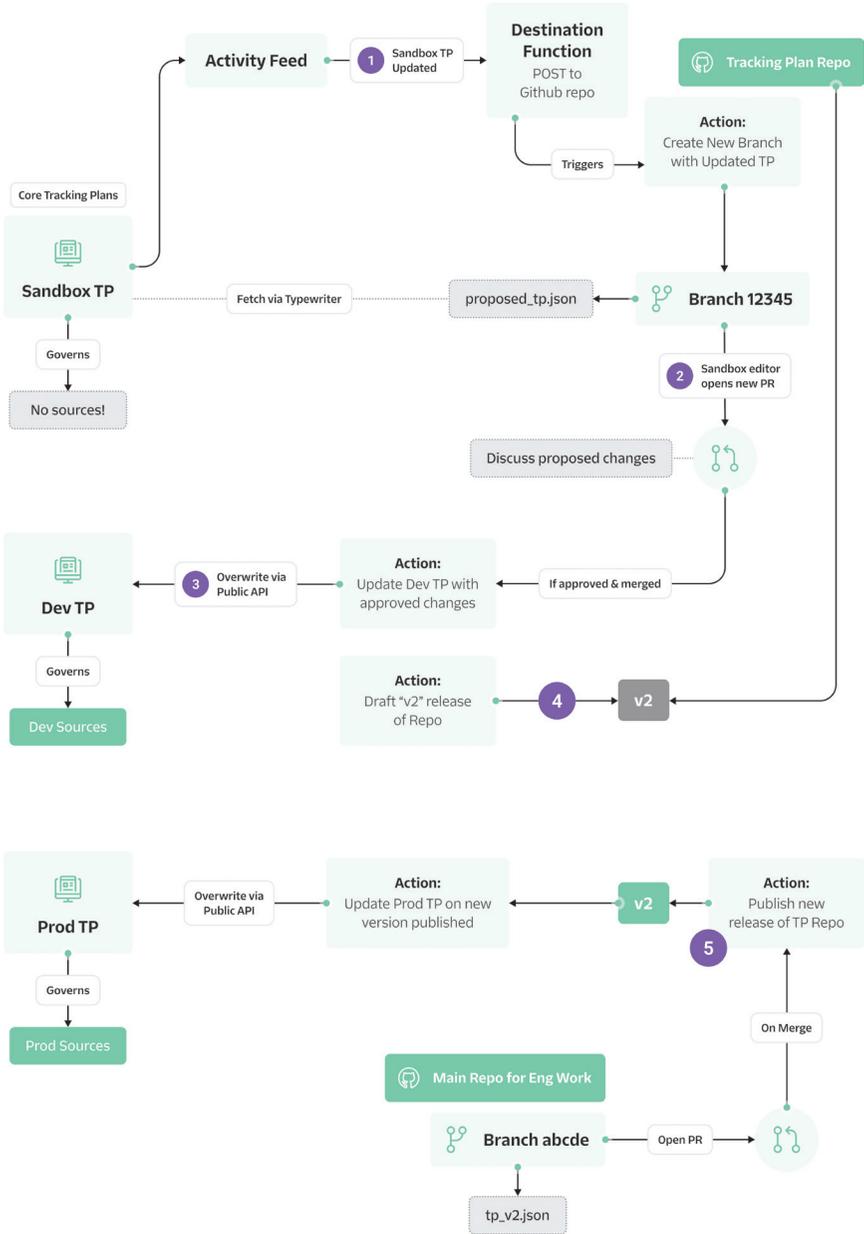
When starting out, it's typical to use a single Tracking Plan for both development and production environments. But as your CDP infrastructure grows and change management becomes paramount, you will need separate environments to experiment with changes to your Tracking Plans without affecting the live data flowing through Segment. Scaling governance necessitates

developing a process to propose changes to data governance so you can safely test and experiment before they're implemented in the production environment.

First, figure out the people component: what will the process be to propose changes to the Tracking Plan, and who will be responsible for approving them? Then figure out your tooling. Will GitHub be the place where changes are proposed via pull requests, and will GitHub Actions power the communication of those changes back to Segment?

With those decisions in place, you can configure automation to interconnect separate Tracking Plans for sandbox, development, and production environments. In the example on the next page we've created three Tracking Plans in Segment (Sandbox, Dev, and Prod) which are connected to branches, merges, and releases in a corresponding Tracking Plan Repo.

Segment tracking plan management architecture



The Sandbox tracking plan is where someone proposes a change to the schema. Once that change is saved in Segment, an Activity Feed event (fig. 1 in the image on page 96) triggers a GitHub action to create a new branch of the Tracking Plan Repo. The Sandbox editor then opens a Pull Request in the TP Repo (fig. 2), which triggers a collaboration and approval process around the proposed changes.

When a change is approved, a GitHub Action pushes those approved changes (fig. 3) from the Sandbox to the Dev TP, which your engineers can use to validate code changes in dev sources. The change approval also creates a draft release of the TP Repo (fig. 4) for use later in the process. Once the engineering work is complete and merged with the main branch of the engineering repo, a final GitHub action (5) publishes a new release of the TP repo, which in turn triggers an update of the Production TP.

This automation flow might seem complex at first glance, but it's an extremely powerful and effective way to manage an ever-evolving set of Tracking Plans in your workspace. Segment's Professional Services team has advised a number of customers on this approach to automated change management in the workspace, who now consider this type of automation critical to their ability to scale data governance for a growing set of stakeholders.

Customer Insight

ENDEAVOR

Industry: Entertainment, Media, and Broadcasting

Endeavor* is a global sports and entertainment company, home to the world's most dynamic and engaging storytellers, brands, live events and experiences.

The company is composed of industry leaders including entertainment agency WME; sports, fashion, events and media company IMG; and premier mixed martial arts organization UFC. The Endeavor network specializes in talent representation, sports operations & advisory, event & experiences management, media production & distribution, experiential marketing and brand licensing.

Endeavor has dozens of users that interact with Segment, many of whom operate within different business units. This includes marketers, product managers, data engineers and scientists, analysts, and members of the IT team.

As these users have discrete roles in the broader organization and often their own distinct business units, relying on role-based access control principles to govern their access makes a lot of sense.

Thankfully, Segment exposes the Segment Public API which enables Endeavor to write code which directly interacts with Segment's identity-access management settings.

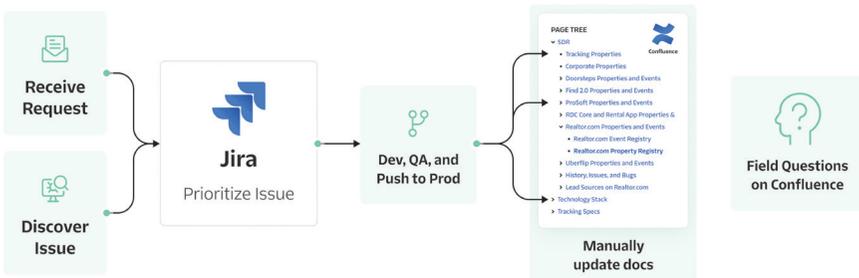
* See QR code across from table of contents

This means the Endeavor team is able to easily reproduce permission structures across Segment workspaces i.e. production and development, and make a full recovery of access management configuration in case of catastrophic failure.

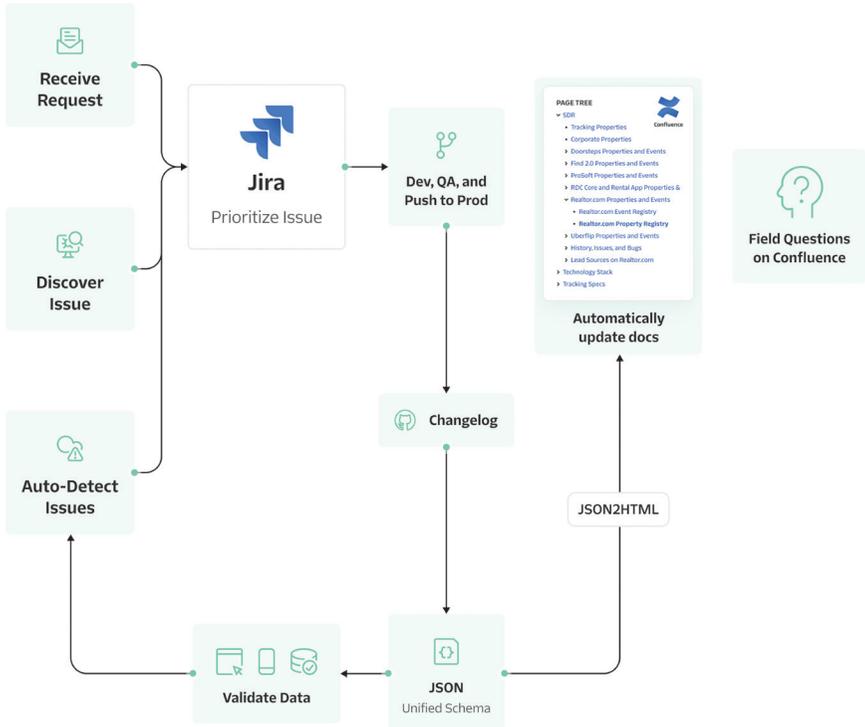
Automated docs

When creating documentation on what to track, what data looks like downstream, etc., most businesses adopt a manual documentation process in a tool like Confluence. In theory, this is so other teams like marketing and product management have up-to-date documentation on how data is flowing through their Segment workspace.

In practice, this creates a complicated workflow for managing changes to that workspace. Let's say a marketer wants to flag an error: they'll email a request to engineering, engineering will prioritize an issue in Jira, they'll push the change to production, and then go back and manually update their documentation to reflect said change.

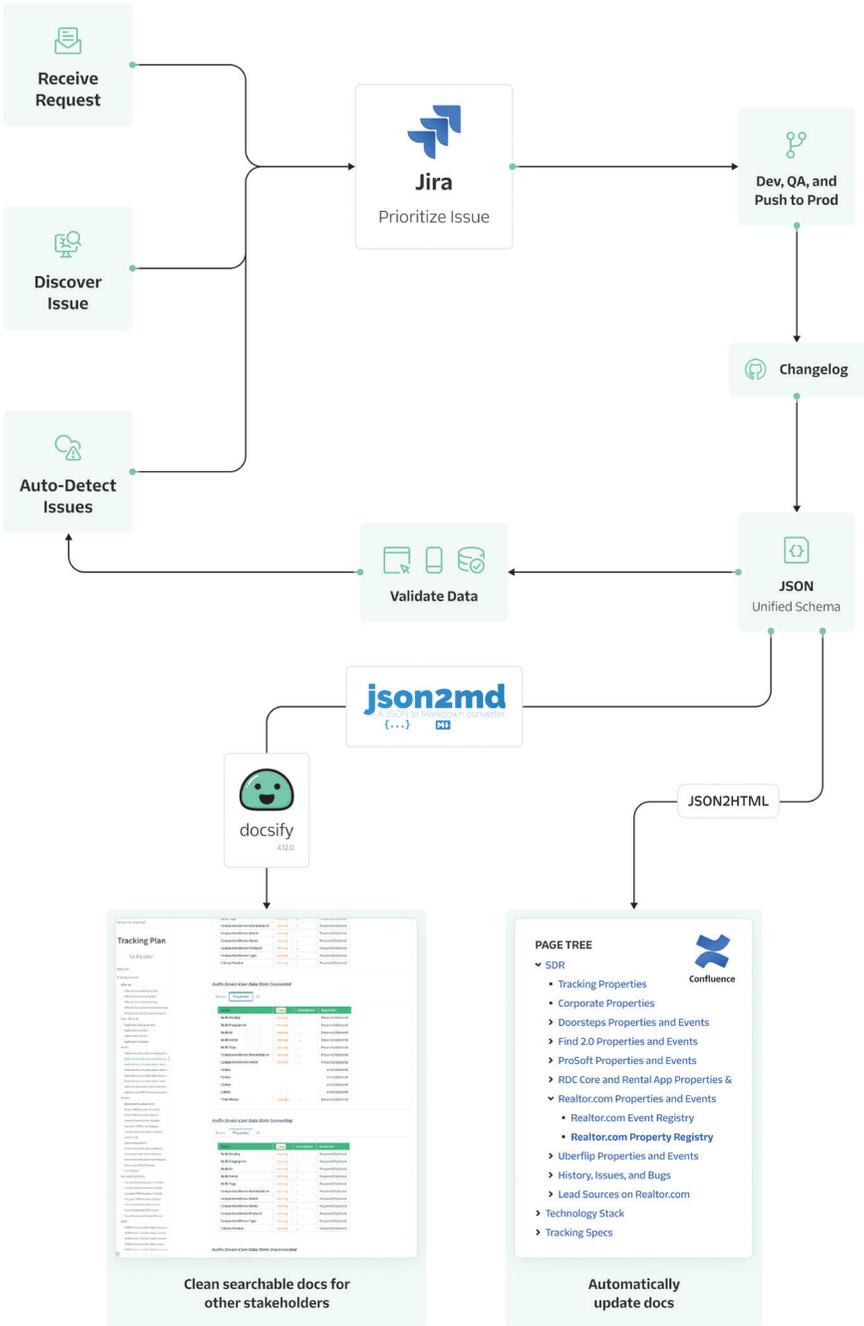


Thankfully, there's a two-part way you can solve this problem. The first is to put a unified schema in place (using a tool such as Segment Protocols) to validate the data and auto-detect issues instead of having the team manually discover errors and emailing engineering to resolve. But you can also take that schema as a json file, update it to HTML, and then automatically add that to Confluence. And voilà! No more manual updates.



But we can also take things one step further.

Many stakeholders within your company want to know less about coding and instrumentation, and more about what the events you're tracking mean within the context of a user journey. For these individuals, you can automatically generate a microsite that contains all of the schema in one place in a super easy, clean, searchable UI. Any stakeholder within the company can access and they can see exactly what everything means inside of Segment.



Scaled SOPs

Once a company has adopted a CDP, a common issue often occurs. As new hires are onboarded, new teams emerge, and as new priorities are introduced, entropy occurs. Those data best practices and frameworks established during CDP implementation? They're quickly abandoned, resulting in a series of silos – the precise problem the business was trying to solve in the first place.

Standard operating procedures (or SOPs for short) help reinforce good data management practices across your organization. We've found that two basic models of ownership can each drive successful adoption of data standards across an organization.



The Wrangler: the sheriff in the wild west of your organization's approach to data management. They help oversee all aspects of your CDP implementation, and are well suited to organizations with a hierarchical structure.



The Champions: this decentralized model places “wranglers” throughout the organization – in product, marketing, engineering, etc. Each champion is responsible for buying into the company-wide data practices, but can also help advocate for their department-specific needs. This model works well for highly collaborative organizations.

Neither of these frameworks is inherently “better” than the other, and their efficacy all depends on the nature of your organizational culture.

Once you’ve identified the best data ownership model for your organization, you’ll need to finalize your SOPs by connecting that model to tools and processes which enable scaled governance and automated documentation. It may seem daunting, but some of Segment’s most innovative customers have done just that.

One such customer established a “Usage Data Council” (also known as a “Schema Council”) of Champions to govern and approve all changes to standardized tracking plans using a GitHub flow similar to the one described earlier, and leveraged a dedicated “Growth Team” to manage all tooling associated with those tracking plans, including automated documentation.

Customer Insight



Industry: Customer Goods & Services

Shipt*, a subsidiary of Target Corp, is a delivery services app offering same-day delivery from a wide range of stores and product categories. Shipt provides personal shopping and delivery services and is available to 80% of U.S. households in more than 5,000 U.S. cities.

Segment Protocols has enabled Shipt to review event tracking and enforce a violation process when plans aren't being followed.

“We went from poorly managed documentation, tracking plans, and event enforcement to a place where people safely and correctly implement new events within Segment” says Andrew Orso, Director of Marketing Data Science, Data & Measurement.

When a team member at Shipt wants to track an event, there is now an enforced process. By utilizing Github to host its tracking plans, teams can reference a comprehensive repository of events tracked, to prevent duplicative events and inconsistencies. Each new event must be submitted to Andrew’s team for review. After a new event is approved, it is synced with Segment Protocols and enforced using the violations feature.

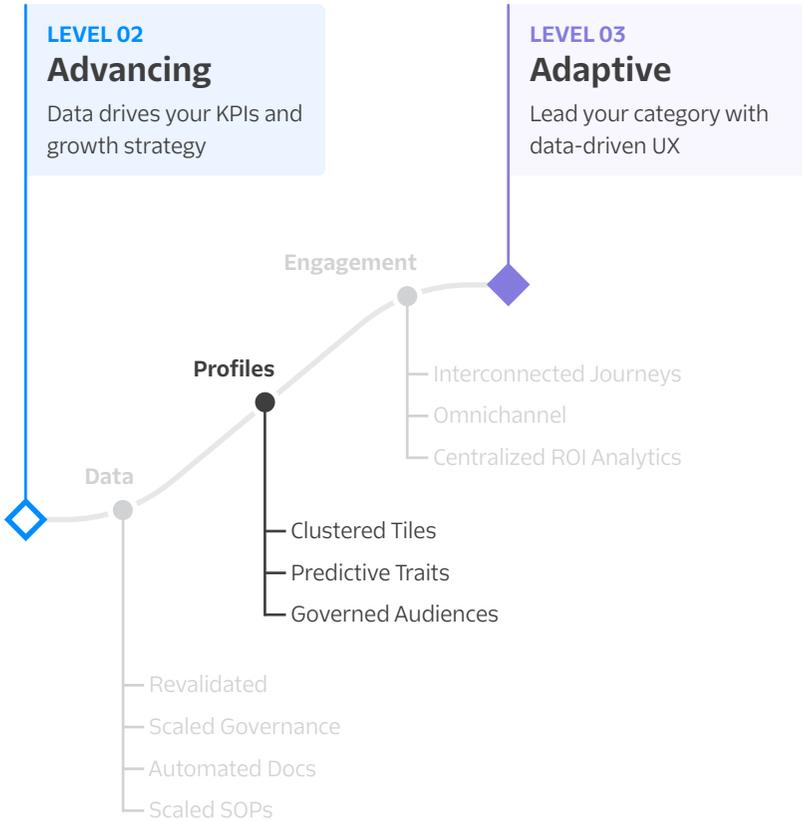
* See QR code across from table of contents

Shipt has now achieved greater consistency in its event tracking with an impressive 3.5B API calls per month (that's 1350 calls per second) and over 400+ distinct events being tracked across 11 tracking plans.

Armed with consistent schema across mobile and web events, Shipt has access to accurate data for cross-channel analytics and activation. As Shipt grows and matures, it continues to iterate on its data governance policies, keeping in mind that it's an ongoing process.

“With Segment, we are helping our cross-functional teams understand how to manage the data, why it's important and ultimately get their buy-in,” says Andrew about the importance of educating teams on data governance.

Adaptive Profiles

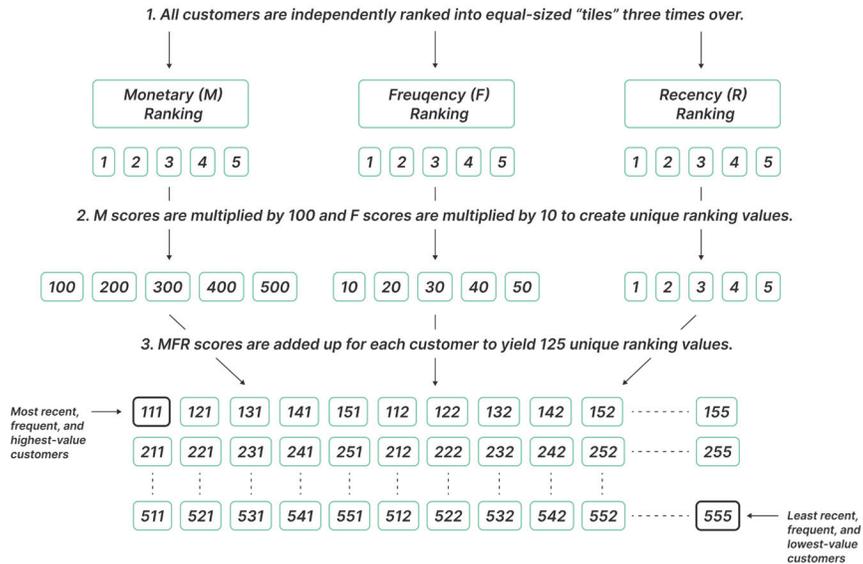


Category leaders don't just collect data, they intelligently leverage data to deliver personalized customer experiences at the right time and place. Getting there requires thinking creatively and strategically about how you cluster and address your users, and how you apply governance principles to your audience-building process.

Clustered tiles

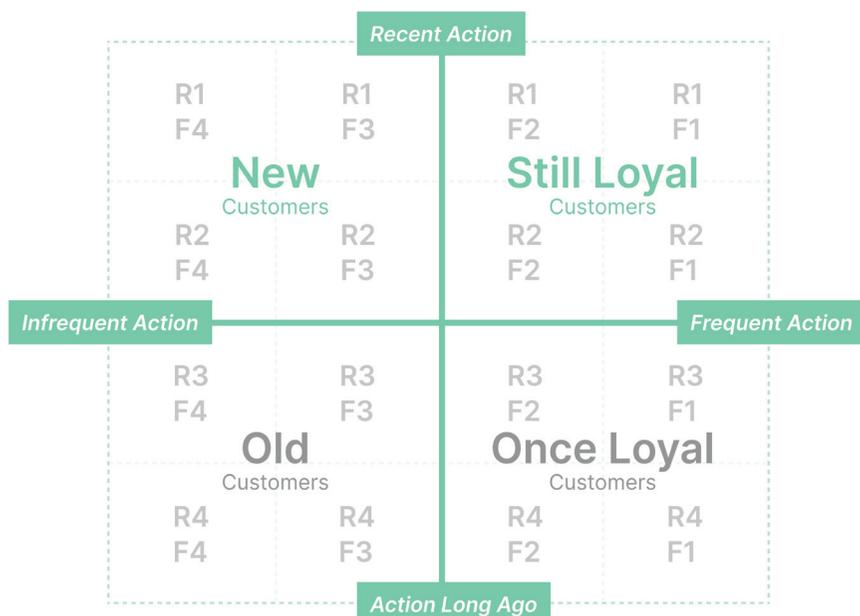
Nearly every business wants to organize its customers into groups based on loyalty (frequency and recency of engagement) and monetary value (LTV or other spend metrics).

Yet businesses often fall into one of two extremes: they either choose hard-coded and arbitrary cut-off values to determine who's loyal or valuable (frequency > 3 orders, recency < 1 year, LTV > \$1000), or they employ incredibly complex micro-segmentation approaches like RFM to generate an overwhelming number of unique segments around Recency, Frequency, and Monetary value:

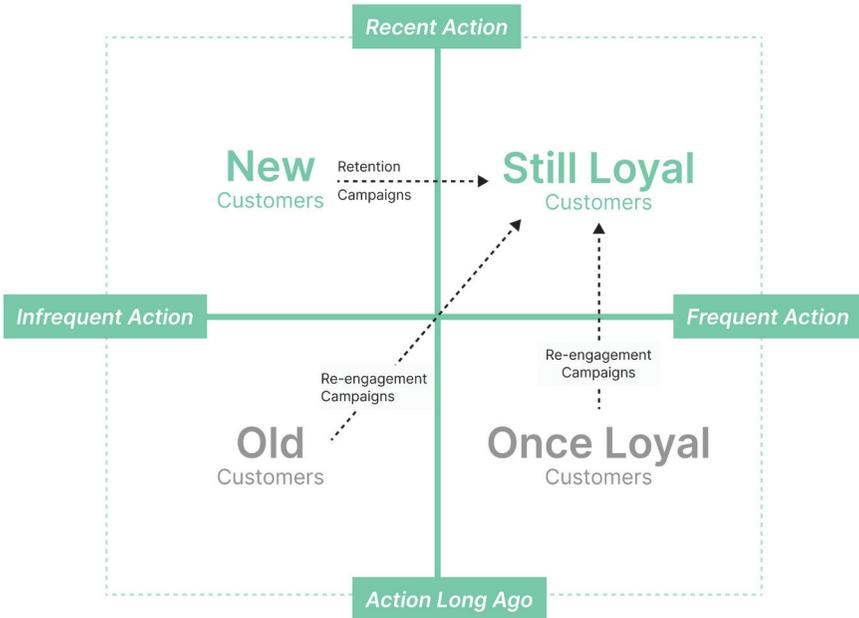


Models like RFM are fantastic for data-driven marketers (and, if run on a regular interval, solve the problem of hard-coded and arbitrary cutoff values for each key metric). But the volume of segments they generate (64 if you're tiling by groups of 4; 125 if you're tiling by groups of 5) aren't very marketer-friendly or intuitive when it comes to building segment-specific campaigns.

That's why we prefer a "clustered" approach to RFM, in which you still "score" your customers around each RFM variable, but you then cluster the outcomes into a much more manageable set of loyalty and value segments:



The result of this clustered approach is a succinct set of intuitive, data-driven segments marketers can use to run well-defined campaigns with equally well-defined target outcomes:

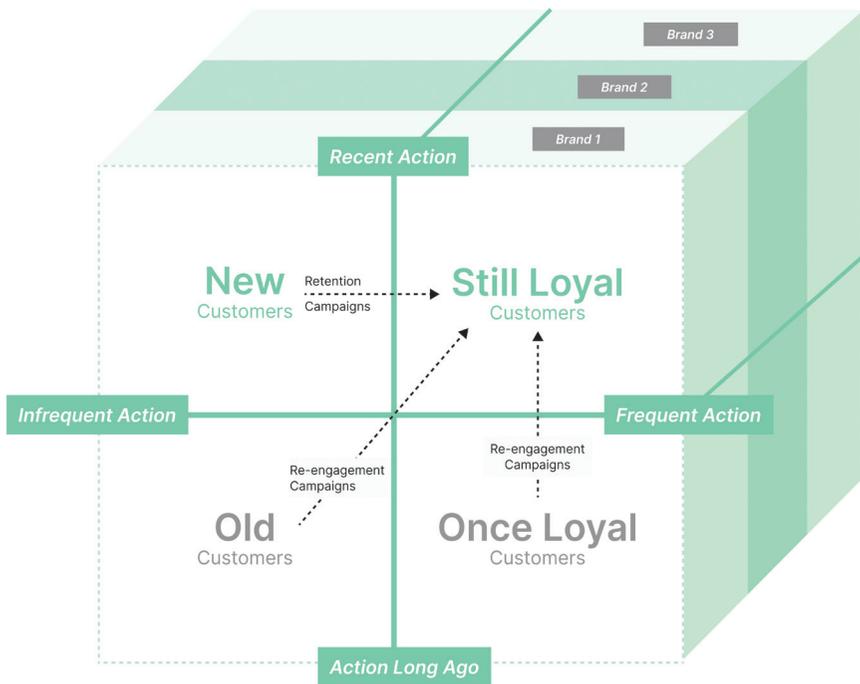


Segment's Personas product (now part of Twilio Engage) has a feature called SQL Traits, which allows you to query your data warehouse, and define audiences with the results. SQL, in turn, has functions that evenly divide returned results into quartiles, quintiles, deciles, and n-tiles.

You get an even split, every time, and because that split is based on the live data from your warehouse, it's always up to date. You build

the model once, and it evolves right along with your product, customer base, and market.

And when combined with other Computed Traits like Most Frequent, you can layer on top of these clustered RFM insights additional affinities related to your brands or products:



Please consult the handy primer blog (Build Quick Customer Audiences with RFM Modeling*) referenced in the additional resources to build an RFM model using a CDP – including SQL code you can plug right into your Segment workspace.

* See QR code across from table of contents

You can also surface the sample code from the blog post inside of our SQL Traits UI, using our SQL templates:

Configure and Preview Your Trait 0

```

1  /*
2  * The following script has been build by Segment Professional Services
3  * to help you generate loyalty and value segments as described on the following blog post:
4  * https://segment.com/blog/rfm-modeling-customer-audiences
5  *
6  * Contact Segment PS if you like to leverage its full potential:
7  * https://segment.com/services
8  *
9  * 1. Extra Metrics:
10 * Note that this will also generate some helpful metrics
11 * Such as LTV (Lifetime value) and AOV (Average Order Value)
12 * as well as loyalty_segment (based of Recency and Frequency)
13 * and value_segment (based on the latter combined with Monetary).
14 *
15 * 2. Adjustments:
16 * a. The number "4" within NTILE yields a 25% split (100/4).
17 * You may consider changing it if you want to break
18 * your population into more or less tiles.
19 * b. The words new, old, still_loyal, once_loyal can also be changed
20 * c. "YOUR_SOURCE" means the schema where your orders table is at.
21 * There's an assumption that your table has a "total" and
22 * a "timestamp" column. You can also change that if required.
23 */
24
25 SELECT
26   anonymous_id,
27   -- stamp underlying R,F,M scores onto profile as traits
28   L.r AS rfm_r,
29   L.f AS rfm_f,
30   L.m AS rfm_m,
31   -- provide useful metrics which drive RFM scores
32   ROUND(CAST(L.ltv AS FLOAT),2) AS ltv.
```

Explore SQL templates ×

All
Segment Data
Payments

Support
Messaging
Sales



RFM-Based Loyalty & Value Segments

Uses a clustered RFM approach developed by Segment Professional Services to generate a loyalty segment (new, old, still loyal, once loyal) and value segment (high value, low value) for any user who has performed Order Completed at least once.

Customer Insight



Industry: Consumer Goods & Services, Retail

Founded in 1960, Domino's* Pizza is the largest pizza company in the world based on retail sales. It ranks among the world's top restaurant brands with a global enterprise of more than 17,200 stores in over 90 markets. With an emphasis on technology innovation, Domino's achieved more than half of all global retail sales in 2019 from digital channels.

Domino's recently launched a series of Facebook and Google Ads campaigns to influence customer conversion and retention. To make these campaigns as effective as possible, Domino's built audiences using Engage (formerly known as Personas) SQL traits with a recency, frequency and, monetary value (RFM) model. Domino's then sent these audiences, which are built directly on top of complete customer data sources, to its ad-tech partners with Twilio Segment to target the right customers at the right time with the right messaging.

With Twilio Segment, Domino's was able to identify 8 cohorts of customers and create personalized customer journeys for each based on behaviors. Domino's now has visibility into which groups of customers are going to buy pizza every 60 days, 90 days, etc., with this

* See QR code across from table of contents

model, which makes ad campaigns through Facebook and Google that much more effective.

By focusing their advertising on high-spending individuals, they saw a 700% increase in ROAS on Google, and a 23% increase in conversions on Facebook.

B2B Consideration: Account-Level Clustered RFM

We can apply Clustered RFM to B2B use cases by means of account-level audiences, enable complex computations on top of accounts – or the combination of accounts and users. So long as you've correctly captured the relevant `account_id` inside of `context.groupId` in an `Order Completed` event, then you can aggregate recency, frequency, and monetary spend across all users associated with an account, and then run a very similar SQL statement to produce account-level traits for loyalty and value.

These traits can then be used to generate account-level coupons, discounts, or other promotions that can be surfaced for any user profile inside that account at their next touchpoint.

Predictive traits

While clustered RFM tiles help you better understand the current state of your user loyalty and value, the next level of data-driven profile building involves *predicting* the likelihood (or propensity) your user might do something in the future.

Predictive analytics is built on the foundation of predicting user intent via lookalike modeling. For example, you predict a customer's

LTV by analyzing customers who exhibit that same behavior (vs. those who don't) and then assign a predicted LTV with a corresponding probability. The same logic works for Next Product to Buy models, which anticipate which product a customer is likely to buy next.

Unfortunately, these projects often require a large lift to operate at scale – efforts which include writing data pipelines, spinning up expensive data science infrastructure, and staffing data science DevOps teams.

Thankfully, there's an easier way to achieve the same result using SQL Traits in Engage. As described in our step-by-step guide*, you can use SQL traits to query Amazon's Sagemaker service to build machine learning models without extensive resources, or needing lots of machine learning experience. If you're looking to take your data a step further, consider using a platform like DBT²⁴ to run transformations in your Data Warehouse, and then use SQL traits to enrich your profiles with the output of those DBT transformations.

Governed audiences

When we talk about data governance, the concept is usually understood in the context of the data that flows in and out of Segment. But as your workspace scales and you leverage more and more capabilities in Engage, you'll want to apply similar governance principles to Computed Traits, Audiences, and Journeys. We call this concept "governed audiences" for simplicity, but it applies to anything else you're building in Engage.

A good governance strategy is built upon a clean, repeatable process and easy automation. To that end, the most basic way to

* See QR code across from table of contents

²⁴ [footnotes.cdmbook.com/24](https://www.f00tn0tes.cdmbook.com/24)

govern audiences is to listen to your workspace. Configure your Activity Feed source to send to a flexible destination like Zapier, and filter down the feed so it only sends Activity events relevant to Audience, Computed Trait, or Journey creation. This can be done with Segment's Destination Filters²⁵ capability.

The screenshot shows the configuration for a destination filter in Segment. At the top, it says "Only Send Event" followed by "to Zapier if...". Below this, there is a section titled "Any" with the text "the following conditions are true". There are three conditions listed, each separated by "or":

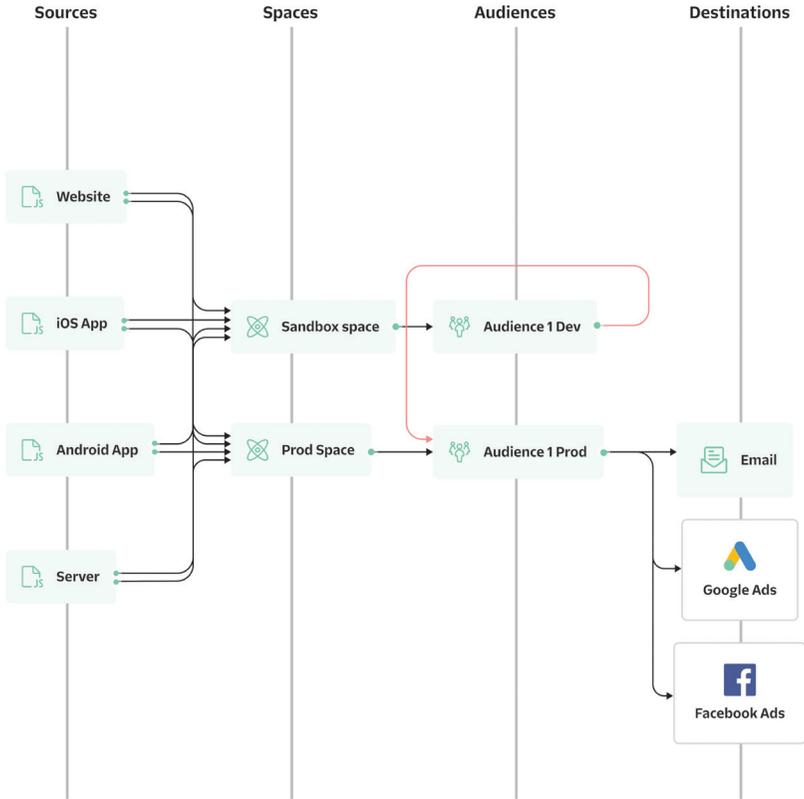
- Condition 1: Event Field is properties.type is (string) Audience Created
- Condition 2: Event Field is properties.type is (string) Computed Trait Created
- Condition 3: Event Field is properties.type is (string) Journey Created

Each condition has a small "x" icon to its right, indicating it can be removed. The conditions are connected by "or" operators.

Beyond just listening, marketers can propose new audiences by mocking them up in a Sandbox Space using the same sources connected to the Prod Space. Then, once you've approved the audience, you can use Segment's clone audiences across spaces feature²⁶ to migrate the new audience to the Prod Space (represented in red on p. 120) without having to recreate the audience definition by hand.

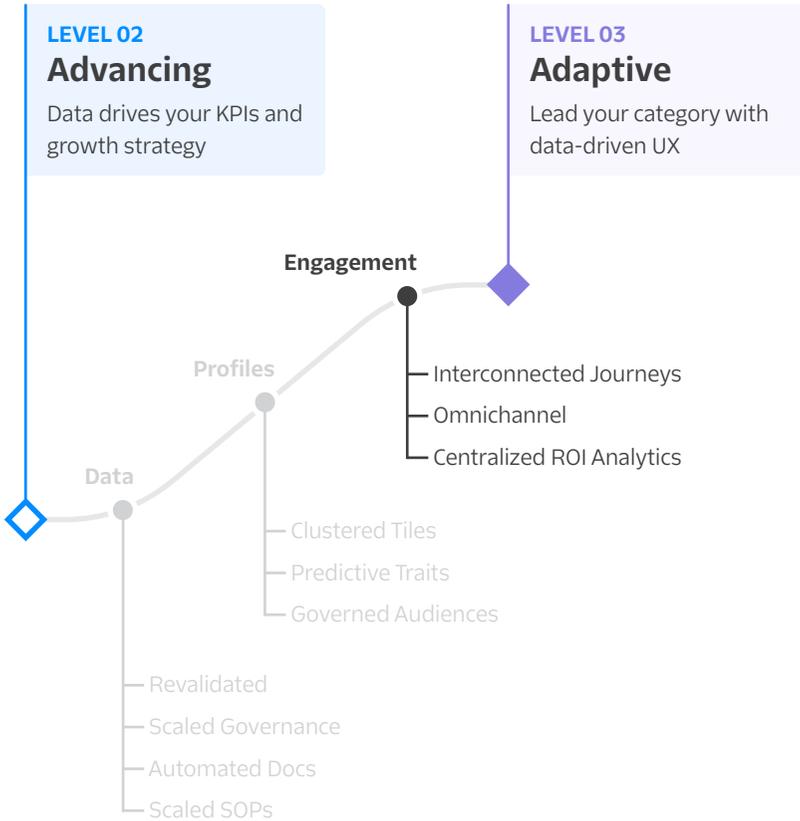
²⁵ footnotes.cdmbook.com/25

²⁶ footnotes.cdmbook.com/26



This will minimize errors while improving the governance of your audiences – and the same exact logic can be applied to Computed Traits. For Journeys, your marketers can “draft” a journey and then request approval before publishing.

Adaptive Engagement



It's no coincidence industry leaders command the position they do – they offer the most comprehensive, cross-channel customer experiences, which raises the bar of what customers expect from competitors. How do you compete in this environment and keep your business growing? By giving your customers plenty of reasons to come back and champion your brand through contextual, interconnected engagement across every channel.

Interconnected journeys

In earlier sections, you learned the nuances of branching logic, and how to engage audiences who have (or haven't met) the specific criteria you've set.

But what happens when even the most sophisticated journey ends successfully? Adaptive-stage customers connect the end of one journey to the beginning of the next, to ensure there's no lapse in customer engagement. This is what we mean by interconnected journeys.

The mechanics of creating interconnected journeys is relatively simple. If one journey is about driving new signups and another journey is about activating those signups, the journeys are interconnected because the entry criteria for the second journey (**Signed Up** X days ago) is tied to the exit criteria for the first journey (**Signed Up**).

The more interesting challenge is ensuring your journeys are interconnected. To that end, make sure you're planning your journeys holistically instead of ad hoc. List out each of the journeys you'd like to build, and their entry and exit criteria. Then ensure the entry and exit criteria line up across journeys. And finally, ensure you're using appropriate time delays within your journeys so you're appropriately pacing engagement. That's the critical thinking and planning you'll need to do so your journeys can drive truly adaptive engagement.

Omnichannel

Omnichannel refers to a business operating on multiple channels, but with an important caveat: these channels are connected so that a consumer could easily switch between them without losing the context of their previous interaction. This ensures the customer experience remains fluid between devices and touchpoints.

We've hinted at some of these capabilities already (see earlier sections such as Multi-Step, and Context Aware) but omnichannel finds itself at the top of the maturity curve because it requires *significant* attention to detail: knowing where the customer is across all those channels, and then delivering them a targeted relevant message whenever they come into one of those channels.

Let's use a fictitious shipping and print services company for an example. A customer goes to the company's website, logs in, and starts looking at photocopy services, printing out a photo, etc. The customer browses for a while, but decides to head in-store instead. The customer has also installed the company's app on their phone, and consented to being tracked for marketing purposes. At the store, the shipping company has established a geofence which triggers once the customer's mobile phone gets within 50 feet of the store. An alert appears on the customer's phone welcoming them, and another alert appears on employee computers at the front counter saying the customer has entered the store. The employee-facing system pings Segment's Profile API and fetches pertinent customer info such as first and last name, as well as what offers they're currently eligible for.

As an employee rings up the customer's shipping order, they're prompted to offer the customer a 20% discount on printing services if

they purchase those services in-store. The customer decides to purchase those services, generating new revenue for the shipping company, and increased loyalty – since the offer for discounted printing was timely and relevant.

Here we had a customer on three channels – web, mobile, and in-store – all pulling from a single unified customer record. And not only were they in sync with one another, they were additive. Each channel passed context on to the next, resulting in a truly differentiated customer experience.

Centralized ROI analytics

How do you measure the success of your marketing campaigns by gauging whether the desired outcomes are produced? In the same way event data from Event Cloud Sources²⁷ can help you target your campaigns more intelligently, you can use the data from Object Cloud Sources to apply attribution modeling to determine the ROI from each action.

Within Object Cloud Sources²⁸, you can track properties such as what campaigns you're running, what your spend cap is, or how many clicks you garnered. By tying these properties back to the UTM (your performance tracking mechanism for the campaign), you can measure the monetary value of each campaign.

This approach becomes most powerful when you sync multiple cloud object sources to your data warehouse (Facebook and Google Ads, for example), and then use some of the tables in those sources to align otherwise obscure identifiers like gclid and fbclid with the query strings already available in your segment data. This helps adaptive CDP

²⁷ footnotes.cdmbook.com/27

²⁸ footnotes.cdmbook.com/28

users unmask otherwise obscured campaign data, and build human-readable attribution models around campaigns.

Oftentimes these attribution models require identity resolution so the specific `anonymous_id` associated with a campaign click or conversion can be tied back to a larger profile associated with a known `user` (or other `anonymous_ids`). Yet performing identity resolution in the warehouse can itself be tremendously tricky, requiring complex logic and costly computations. For this reason, Segment enables its customers to sync existing identity resolution into the warehouse using a capability called Profile Sync²⁷, which appends all events and identify calls associated with an Engage Space with a canonical `segment_id`, to make attribution modeling much more efficient and cost-effective.

As you build those attribution models, remember to consider the role each ad click plays in the overall engagement journey. Even if a campaign didn't drive a conversion, that doesn't necessarily mean the channel used for that campaign wasn't valuable. You'll likely find that some channels are better at driving initial discovery or consideration while others help explicitly drive the sale, and so each should be considered for its strength before your next omnichannel campaign.

²⁹ footnotes.cdmbook.com/29

Assessing Adaptive Success

Reaching the Adaptive stage of data maturity is a tremendous accomplishment only achievable by industry-forward organizations ready to act as a category leader in their sector. At this point, your CDP implementation will enable you to actively anticipate your customers' needs and to create contextual, relevant engagement across channels: a differentiator that not every competitor will be capable of employing. Here's a brief recap on what it takes to reach this point:

- **Adaptive Data** is revalidated regularly, uses automation and well-defined SOPs to scale governance, and is automatically documented in different formats for different stakeholders.
- **Adaptive Profiles** leverage clustered tiles and predictive traits to generate highly targeted engagement, and are built around a governance process for audiences, computed traits, and journeys.
- **Adaptive Engagement** interconnects journeys and leverages true omnichannel messaging to optimize engagement, and takes advantage of cloud object sources to compute the true ROI of your marketing campaigns.

It's very important to stress that your journey with Segment doesn't end with reaching the top level of data maturity. Your data strategy is a continuous, iterative process, and in each stage of data maturity, we focus not only on the initial implementation but also developing sustainable practices that are meant to be employed in the long term.

Case Study



How Allergan built an adaptive DTC strategy using Twilio Segment

Allergan's* legacy loyalty program was impersonal and unengaging. Fragmented data led to a generic, one-size-fits all legacy loyalty program app, Brilliant Distinctions. Half of the users in the program were added by their doctors and didn't know they were signed up. With only 2.6 out of 5 stars on the App Store, the team knew they needed to improve the user experience and engage with customers in more personal ways.

That's when the Allergan team turned to Twilio Segment.

Using Segment, the Allergan team could capture and standardize customer data across Allergan's portfolio of brands, and automatically load that data into Allergan's data warehouse, Snowflake.

Using the Segment-enriched data in Snowflake, the product engineering team built out machine learning models to predict relevant offers, products, and content for each customer. They pipe these predictions into their applications and websites to deliver personalized content and increase engagement.

Armed with centralized, trusted customer data through Segment, plus machine learning models and AI-based audiences built

* See QR code across from table of contents

with Segment and Snowflake together, the team is able to deliver customized content, offers, and rewards directly to customers through Allē.

For example, Allergan leverages Personas (now part of Twilio Engage) to identify what types of content visitors engage most with in order to recommend the most relevant treatments. Similarly, the team is able to create audiences of users that searched for providers, but have not used any offers, then send unique coupons to those interested groups. Timely, relevant, and direct communication with customers accelerates product cross-sells, encourages repeat visits, and enhances the user's journey across Allergan's portfolio of products.

With Segment at the heart of their data architecture, Allergan's new DTC program resulted in \$400M+ in sales in 2021.

Conclusion

Whether you're just starting out on your customer data journey, or you're in need of a hygiene check on a data strategy already in flight, there are a number of recommendations we'd like to leave you with.

- **Clearly assess your current customer data capabilities.** Whether you're plotting the course of your customer data strategy or performing a checkup on a strategy you've been operationalizing, understanding where your company ranks against best practices can help identify key areas for focus and improvement. Perform a detailed diagnostic of your company's current practice to identify the key steps in your company's roadmap to becoming an insights-driven business.
- **Customer data is a team sport.** Customer data maturity is predicated on organizational collaboration, and can't be dependent on just one department. It requires individuals across your organization who can evolve the standards as your customer data strategy matures.
- **Understand this work is never really done.** It's a big lift to establish your company's customer data strategy, so why not maximize the lifespan and value of this work by building in continuous optimization as the final step of your roadmap.

Our hope is that this guide has helped you establish a clear-eyed assessment of your current capabilities and needs.

Looking Ahead

What if every marketer, call center agent, and sales rep had a clear understanding of each and every customer they interact with? Who they are, what makes them unique, and what they care about.

What if they knew how each interaction would positively or negatively impact the experience? What if they were capable of acting on this on an individual level, at global scale?

Thanks to the rise of AI, these are no longer questions of “if”. They’re questions of “when”.

AI has leveled the playing field for customer engagement; most businesses have the tools to compete effectively, and you no longer need a team of engineers or data scientists to gain valuable insights from AI.

In the next ten years (and let’s face it, probably even sooner), these real-time, context-rich experiences, once the preserve of the most technologically advanced companies in Silicon Valley, will be available to companies across the globe, big or small.

While it’s impossible to predict how AI will continue to evolve in the next few years, what we know for certain is that it will require a mature approach to customer data.

If your customer data within your organization is siloed, inconsistent, stale, or incomplete, even the most innovative AI applications will fall at the first hurdle.

That’s why it’s never been more important to follow the customer data maturity framework contained within this book. Not only will it help you lay the groundwork for AI in your organization today, you’ll learn how to refine your data strategy to any future tectonic shifts in customer engagement coming down the line.

About the Author

As the leader of Segment's Advisory practice at Twilio Segment, Seth Familian guides Segment's largest customers on data architecture strategy, growth marketing innovation, and data governance optimization, with Twilio Segment at the core.

After ten years of running his own data strategy consultancy, and five years of performing Segment-first strategy and integration oversight work in consulting engagements, Seth joined Segment in 2018 as a seasoned Solutions Architect.

He quickly gravitated towards large, complex, strategy-focused projects, which led him to found the Advisory Services practice in 2020. Seth and his wife live in New York.



Appendix

- 1 Segment Protocols: footnotes.cdmbook.com/1
- 2 GitHub Scroll Depth: footnotes.cdmbook.com/2
- 3 GitHub Consent Manager: footnotes.cdmbook.com/3
- 4 The Group Call: footnotes.cdmbook.com/4
- 5 Account-level audiences: footnotes.cdmbook.com/5
- 6 Privacy Portal: footnotes.cdmbook.com/6
- 7 Labels in Segment: footnotes.cdmbook.com/7
- 8 Event and property libraries: footnotes.cdmbook.com/8
- 9 Functions: footnotes.cdmbook.com/9
- 10 Transformations: footnotes.cdmbook.com/10
- 11 Audit trail: footnotes.cdmbook.com/11
- 12 Event delivery: footnotes.cdmbook.com/12
- 13 Replays: footnotes.cdmbook.com/13
- 14 Reverse ETL: footnotes.cdmbook.com/14
- 15 Historical data: footnotes.cdmbook.com/15
- 16 GitHub CSV Uploader: footnotes.cdmbook.com/16
- 17 Amazon S3 Source: footnotes.cdmbook.com/17
- 18 Analytics integrations: footnotes.cdmbook.com/18
- 19 Destination methods: footnotes.cdmbook.com/19
- 20 Reconcile anonymous visitor data: footnotes.cdmbook.com/20
- 21 Querystring API: footnotes.cdmbook.com/21
- 22 Segment Computed Traits: footnotes.cdmbook.com/22

- 23** Segment Profile API: footnotes.cdmbook.com/23
- 24** DBT: footnotes.cdmbook.com/24
- 25** Destination Filters: footnotes.cdmbook.com/25
- 26** Clone audiences: footnotes.cdmbook.com/26
- 27** Event Cloud Sources: footnotes.cdmbook.com/27
- 28** Object Cloud Sources: footnotes.cdmbook.com/28
- 29** Segment Profile Sync: footnotes.cdmbook.com/29

Additional Resources

To help guide you through this book, we've collected a series of resources to aid you on your customer data journey. For links to these resources, please use the QR code opposite the Table of Contents or visit segment.com/customer-data-maturity-resources

Introduction

The 4-Step Guide to CDP Success (Guide)

What Does It Take to Achieve Good Enterprise Data at Scale? (Blog)

Foundational Maturity

Goals & Use-Cases activity (Segment Jumpstart Lite program)

Choosing Metrics that Matter (Analytics Academy)

How IBM scaled its customer data foundation across 150 product lines and increased revenue with Twilio Segment (Case study)

First-Party Data: How to Activate & Leverage Your Data (Blog)

Streamline Compliance with Segment's End-User Privacy Tools (Recipe)

Identity resolution: the definitive guide (Blog)

How iconic guitar maker Fender® increases customer engagement and lifetime value with its Fender Play® music lessons app (Case study)

Identity Resolution: A Guide to The Post-Cookie World (Guide)

How Toggle® speeds into new markets with connected data and tools (Case study)

Advancing Maturity

Personalize your Twilio IVR with Segment (Blog)

Adaptive Maturity

How Endeavor used Twilio Segment to scale its data infrastructure (Case study)

Build Quick Customer Audiences with RFM Modeling (Blog)

How Domino's decreased cost per acquisition by 65% with Twilio Segment (Case study)

Run predictive analytics with SQL commands (Recipe)

How Allergan Aesthetics relaunched its Allē loyalty program and generated \$400 million in direct-to-consumer sales (Case study)

