



SNOWPLOW

POWERING YOUR DATA JOURNEY

HOW TO DESIGN A TRACKING PLAN THAT WILL MEET THE UNIQUE DATA REQUIREMENTS OF YOUR BUSINESS

A step-by-step guide to creating and implementing a centralized tracking plan that enables organizations to work with high quality and consistent data sets.

[SNOWPLOWANALYTICS.COM](https://snowplowanalytics.com)

SPEEDREAD:

DESIGNING A TRACKING PLAN IN 60 SECONDS

- Designing a tracking plan starts by gathering the data needs of different functions within the organization, and finding commonalities and areas of overlap.
- A carefully considered tracking plan will define what event data is tracked, how data is structured, and who takes ownership of the data strategy.
- Common challenges around data meaning and structure can be overcome using schemas, helping you to define the structure of behavioral data and apply one common language to define data across the business.
- Using schemas will allow for a highly expected data structure that can easily evolve over time. Modular tracking enables data teams to create a set of core schemas that can be easily evolved to serve multiple data needs.
- Tracking design is personal to each organization. We'll share some best-practice advice and examples, but every business should build their tracking plan around their core use cases.

The need to empower your business with accurate data collection

It's no secret that businesses in 2020 rely heavily on data in order to stay ahead and succeed. A recent [Gartner report explains the importance of meaningful data](#), stating: "By 2023, data literacy will become an explicit and necessary driver of business value, demonstrated by its formal inclusion in over 80% of data and analytics strategies".

However, there is a risk of organizations focusing too much on how to use their data and the tools they have to visualize insights, and not enough on the actual data collection itself. In the race to become data-informed, companies may be getting ahead of themselves and not paying close enough attention to the foundation of any data strategy: data collection that enables them to collect behavioral data that is reliable, easy to understand, and easy to work with.

Designing and implementing an effective tracking plan is a prerequisite for a powerful data strategy. It defines what data an organization wants to collect, and how that data is then structured and formatted to allow data consumers to analyze it.

Here at Snowplow, we recognize just how powerful data collection is as a business tool. We're always happy to share best practice thinking to empower your business to collect the data that's relevant to your use cases.

Our guide to designing a tracking plan looks at what a tracking plan is, the concepts behind successful data capture and why it's vital for your business. We'll also address some of the challenges that can arise when working with different teams who each have different data needs.

Why you need a good tracking plan

A tracking plan is a live resource document, often in Excel, that outlines a plan to accurately capture and collect event data, such as customer journeys on a website or web platform.

There are typically three components that are tracked for any business:



An event and the properties needed to track them



Where to track them (what triggers the event)



The reason why your business wants or needs to track them

A carefully considered tracking plan can help you to create a core data capability within your organization, which will unlock a number of benefits for your business, such as:

- **Understanding the value of marketing campaigns**
- **Understanding user intent**
- **Optimizing your customer journey**
- **Evolving the business and products to meet customer demands**

Overcoming common data challenges

A tracking plan can help your business overcome some of the more common challenges that crop up with data collection. Here are some of the obstacles that can often prevent organizations from utilizing the full value of their data.

Data formatting issues

Data captured from a range of digital products, platforms, and devices is formatted differently. This can make it harder to analyze your captured data as there may be different naming formats or referencing. A good tracking plan will help eliminate this issue and keep your data high quality and with consistent naming conventions across the board.

Differences in data structures

There may be teams of people who have different requirements, and therefore different preferred data structures. For example, analysts may desire highly structured data that makes it much easier and faster to analyze. Front-end developers, on the other hand, may find it easier to implement tracking for unstructured data. Having a central tracking plan will ensure consistency across all teams and platforms, so e.g. your mobile and web data will be structured the same regardless of source.

Data governance

Multiple teams with various data needs and therefore different tracking designs can make data governance problematic too. Each team may operate separately from colleagues in other functions and only focus on what's important to them. The result of this is that there is no central data capability, and no one responsible for managing the data needs of the business as a whole. Having a team solely responsible for owning data capture ensures there's a consistent approach across the organization.

Data meaning: now, and in the future

A lack of effective data governance can result in sprawling event dictionaries that should define the meaning of the data being collected. For example, there may be multiple versions of spreadsheets, each with different definitions, which will dilute data meaning and impact what the business is able to do with that data. Similarly, your products and use cases will evolve over time, meaning your event dictionary will need to evolve with it. A tracking design strategy update needs to be carefully governed to help modifications and additions go smoothly.

Questions to ask before you get started

Before designing a tracking plan, it's worth your organization considering a number of more in-depth questions so you can successfully design a tracking plan that best suits your requirements. These are not questions that one person alone can answer; we recommend that all senior decision makers and heads of functions have an input to ensure their needs are documented and understood.

What behavioral data does our business need?

Every business is unique, which makes data collection a bespoke activity. Likewise, each team within your organization will understandably have a greater focus on data relating to their own use cases.

It's important to conduct company-wide analysis to understand requirements and identify overlapping needs.

What are your key objectives across the business when it comes to collecting behavioral data? Are these being covered in the plan design?

What are the needs of the team collecting the data?

In short, does the tracking plan give the data team the tools they need to collect the behavioral data demanded by the various teams and data analysts?

What are the needs of data consumers?

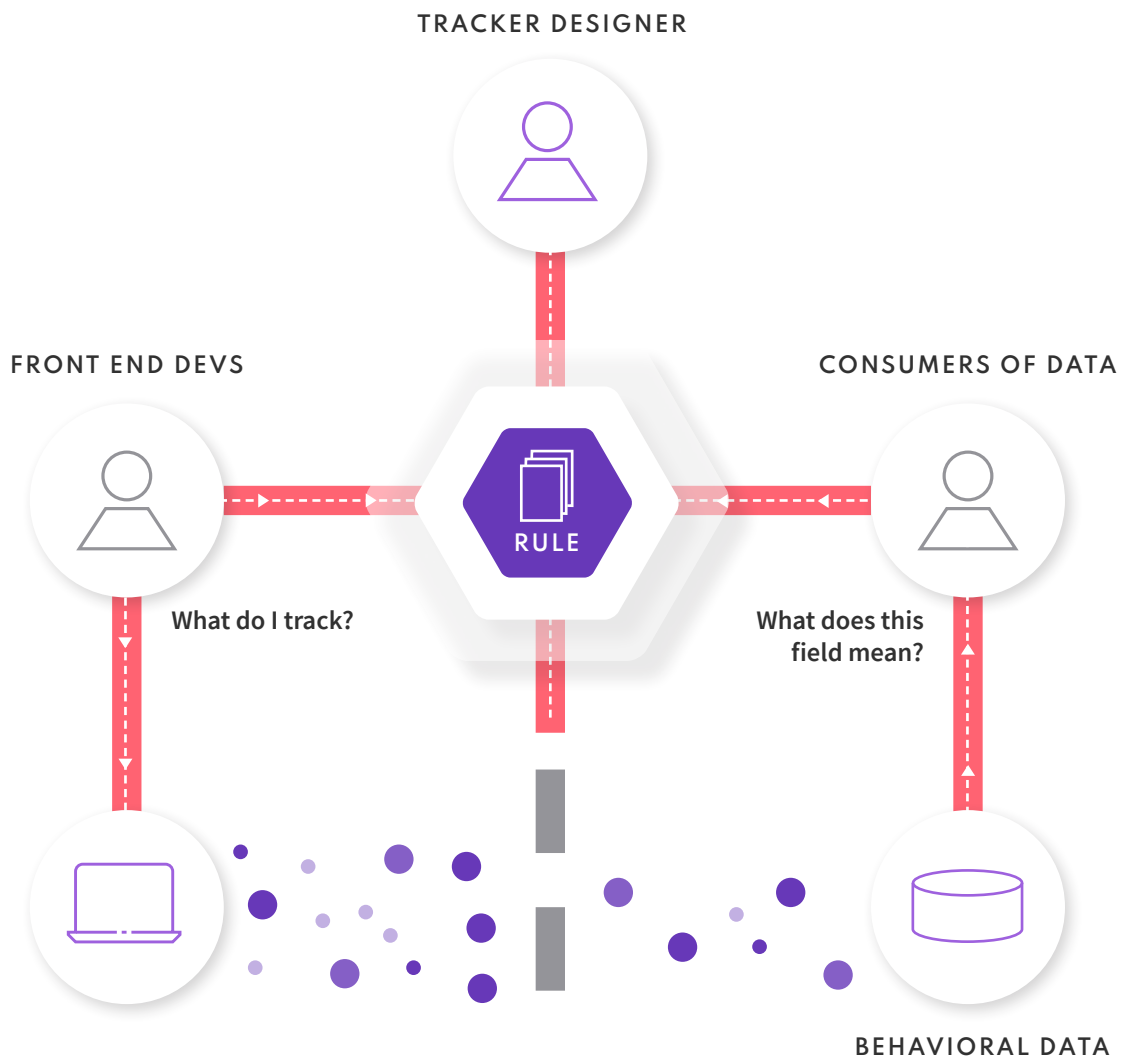
The old cliché says that analysts spend 80% of their time cleaning and formatting data, and only 20% of their time analyzing it. Does your tracking plan deliver clean, accurate, and structured data so that analysts can start deriving value from it immediately?

Do we have a ‘gatekeeper’ acting as a single point of reference within our organization?

Assigning a gatekeeper can help improve your data collection processes in a number of ways. A gatekeeper will typically:

- **Consolidate every team’s requirements and identify commonalities to help to avoid multiple versions of the same thing. For example, different teams may all need to know the identity of the user, but each could have slightly different names of the same thing: ‘user name, username, user_name’ etc. A gatekeeper will help create one centralized version.**

- > Manage stakeholder needs and expectations, and be in the best position to work collaboratively across different teams and business functions.
- > Build a hierarchy of the data that needs to be captured so there is a central source of 'truth' for the business. This creates a far more efficient and accurate approach to data meaning, providing one reliable source of information that can be shared across the organization.



Tackling these challenges with schemas

The great news is these challenges can be overcome by using schemas. Schemas define the structure of the behavioral data that you collect, and enable you to apply one common language to define data across the business. Benefits of using schemas include:

- **Combining the flexibility of an unstructured approach with the predictability of structured events.** You have the freedom to capture the data that is most important to your business, while also giving it a degree of structure and making it clear what that data should look like.
- **Capturing data meaning and intent.** Your team can use descriptions to make sure everyone understands what the data means from a tracking and analytical perspective.
- **Enforcing data governance requirements.** You can control and version schemas to understand when and how they evolve.
- **The ability to evolve with your business.** They are explicit in determining what data is tracked, so you can make changes as required to meet the business needs.

How schemas are designed are important to a data collection strategy. An ideal scenario is one where schemas are flexible enough to evolve over time as the business needs and products change, but simultaneously sturdy enough to still give understandable meaning and ensure data quality.

Modular tracking for greater efficiency and accuracy

Modular tracking means creating and maintaining a core set of schemas and combining them in a modular way to capture all the behavioral data your organization needs.

You can use one schema that can be leveraged multiple times to suit individual needs. This is done to avoid having to create multiple schemas that are difficult to manage, can get very messy, and result in the sprawling dictionary problem referenced earlier in this paper.



Group events	Combine entities				
signup_flow	screen	user			
checkout_flow	screen	user	product	product	
engagement	screen	user	product		
search_update	screen	user	product	product	filter

Top tip from our team: limit the total number of schemas applied to your site to between 10 and 20.

Snowplow's approach to data capture and structure

The way we structure events and entities is one of Snowplow's most unique features. Events and entities describe your users' behavior and actions. An event is an action at a point in time, such as a click, pageview, purchase, or swipe. Entities give context to an event, such as user ID, subscriber, or location. Entities typically describe the environment in which the event happened.

For example, these are the things we want to track:



And here's how this information is grouped:

The diagram illustrates the categorization of properties for Events and Entities. It is titled "WHERE DO THE PROPERTIES SIT" and features a hexagonal icon. The content is organized into two main columns: "Event" and "Entity".

Event		Entity
Default	Custom	Custom
user_id	button_name	content_title
timestamp		author
page_url		date_created
mkt_camp		content_id
device		
city		

Both events and entities can have properties, and they are both defined by schemas. Schemas are important at every stage of the Snowplow pipeline as they:

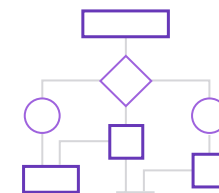
- > **Define what events need to be tracked and collected**
- > **Determine how data is validated after collection**
- > **Define the enrichments that enhance event data**
- > **Define the data structure when the data is loaded into the data warehouse**

Examples of design tracking

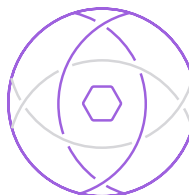
At Snowplow, we see our customers design their tracking to serve a range of use cases. Here are four common examples that highlight how flexible and customizable tracking can be:



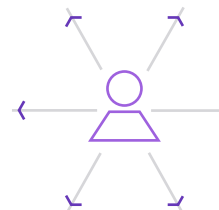
Tracking events based on actions:
such as clicks, views, conversions,
and video plays.



Tracking events based on functionality: such as sign-up flows
or checkout flows.



Tracking events based on objects:
whether the user has viewed a
module, interacted with a feature, or
returned to a specific part of a page.



Tracking events based on organizational structure: whether
the user has interacted with different
teams or functions, such as HR for
recruitment and careers, marketing,
or product.

Let's take a look at some live tracking plans for various use cases.

USE CASE 1:

Tracking scroll

In this example we want to understand a number of factors, such as which suggested posts are popular, and how far users navigate through an article.

- Capture basic data around device type, location, user, title of the post being viewed etc.
- Each row captures scroll progress, including how many posts have been displayed, and the titles and authors of those articles.
- Each set of displayed posts fires an event notification to track what is being viewed.
- Track if the user scrolls back up to view a post displayed earlier in the page.
- Once an article is clicked to view, track scroll as a percentage or within content modules as the user navigates the page: eg, an event is fired if 70% of the post is in view.
- Track changes from mobile to laptop or mobile to app view if the user swaps device.
- Understand the most engaging content, identify pages and assets that are popular and which ones have a higher drop off rate. This allows you to make content recommendations and optimize content production.



EXAMPLE USER ACTIVITY: POST VIEWS ONLY

Timestamp	Post position	Subscribed user	Screen name	Post name
2020-10-10 10:00:00	1/4	TRUE	homescreen	a123
2020-10-10 10:00:13	2/4	TRUE	homescreen	b123
2020-10-10 10:00:55	3/4	TRUE	homescreen	c234
2020-10-10 10:00:57	4/8	TRUE	homescreen	d435
2020-10-10 10:00:58	5/8	TRUE	homescreen	e123
2020-10-10 10:01:45	6/8	TRUE	homescreen	e134
2020-10-10 10:01:59	7/8	TRUE	homescreen	b342
2020-10-10 10:04:01	6/8	TRUE	homescreen	e134
2020-10-10 10:04:10	5/8	TRUE	homescreen	e123

USE CASE 2:

A sign-up funnel

In this example we want to understand how users progress through a sign-up process, and identify the cause of drop-off or sign-up failure.

- **Track progress through each step, and log them as separate events.**
 - Screen view tracking sends an event notification when a page is loaded.
 - Each event is sent with a screen name entity showing which page the user is on.
 - Each step has a value and failure reason properties, to show which step the user is on, what values they've inputted during that step, and the reason for sign-up failure (where relevant).
- **Understand how easy it is to go through the sign-up process.**
- **Track how long each step takes, and identify common abandonment points to improve the user experience**



EXAMPLE USER ACTIVITY: SIGNUP FLOW

Timestamp	Event name	Signup step	Step value	Failure reason	Screen name
2020-10-10 10:00:00	Screen View				Login Signup
2020-10-10 10:00:13	Signup Flow	Create account			Login Signup
2020-10-10 10:00:04	Screen View				Enter Name
2020-10-10 10:01:18	Signup Flow	Enter first name	tracked_by		Enter Name
2020-10-10 10:01:22	Signup Flow	Enter first name	snowplow		Enter Name
2020-10-10 10:01:30	Screen View				Enter Date Gender
2020-10-10 10:01:49	Signup Flow	Enter birthday	1977-05-04		Enter Date Gender
2020-10-10 10:01:53	Signup Flow	Enter gender	rather_not_say		Enter Date Gender
2020-10-10 10:02:14	Screen View				Enter Email
2020-10-10 10:01:49	Signup Flow	Enter email	tracked_snowplow	Invalid_character	Enter Email
2020-10-10 10:01:55	App background				

DROP OFF

USE CASE 3:

Search behavior

In this example we want to understand what terms users are searching for, the results they're shown, and their progress to successful purchase at checkout.

- Track the keywords being searched and how many results were returned.
- Identify the user and determine if they are registered or a subscriber.
- Track if the user logs in, and what we know about them (such as age and location).
- Track each event and activity through to checkout.
- What is the total cost of the purchase? Does it match the unit price of the item, or has a discount code or coupon been used?
- The retailer can then understand who is searching, what they're searching for, what results they are shown, whether they converted, or where they abandoned. They can then build segmentations around demographics, develop features such as "you might also like", and optimize the checkout flow to increase conversion rates.



event_name	user_id	timestamp	search_event	checkout_event	product_entity	user_entity
search_update	abc123	2020-10-10 11.42.40	{ "keywords": "donkey_kong", "num_results": 13941 }		{ "sku": "1231u13", }, { "sku": "5312k52" } }	{ "subscriber":TRUE, "gender": "f", "age": 34 }
checkout_flow	abc123	2020-10-10 11.42.40		{ "step": "add_to_cart" }	{ "sku": "1341u13" "price": 7.99 }	{ "subscriber":TRUE, "gender": "f", "age": 34 }
checkout_flow	abc123	2020-10-10 12.12.10		{ "step": "transaction" "coupon": "APRIL20" "basket_value": 6.20 "items": 1 }	{ "sku": "1341u13" "price": 7.99 }	{ "subscriber":TRUE, "gender": "f", "age": 34 }



Start your data collection journey now

Designing and creating a tracking plan will give you greater control of your data collection strategy. It will help to define the data you collect, what your data means to your business, and how your data is structured.

By using a universal language (schemas) to determine a common understanding of the data meaning

and what it defines, your business will ensure everyone is on the same page.

By adopting a modular approach, you can design a tracking plan that not only provides you with rich and accurate data to solve today's use cases, but is also flexible enough to deliver against future requirements.

Talk to Snowplow today to discuss your tracking design requirements and learn how to get started. Get in touch to book a personal demo.

[Book a demo](#)