Data Mesh for Event Based Data

Recipe to mature data cultures from wild west to data quality at scale

Table of Contents

Abstract

As organizations scale, the challenge of maintaining high-quality, accessible, and reliable event data intensifies. Data governance frameworks such as Data Mesh and Data Contracts offer solutions, but many companies struggle to implement them effectively. This white paper provides a tactical guide to achieve event data quality at scale, transforming chaotic, ad-hoc data management into a structured, automated system that enables both speed and quality.

We explore the stages of data culture maturity, from the Wild West of fragmented data collection to centralized governance bottlenecks—and ultimately to a scalable self-serve model. Through real-world case studies, we demonstrate how leading companies have implemented federated data governance, automated data validation, and systematic schema management to accelerate analytics while ensuring data reliability.

This paper outlines a recipe with the key ingredients and secret sauce required to operationalize Data Mesh and Data Contracts, providing a roadmap for organizations to eliminate manual enforcement, empower domain teams, and build trust in their data—without slowing down innovation.

Introduction

Data quality is a major barrier to scaling—60% of tech executives cite it as their biggest challenge (1 - McKinsey). The impact is costly: data scientists spend 45% of their time cleaning data instead of generating insights (2 - HFS Research), while poor governance and inconsistent taxonomies create misalignment across teams.

Investing in data quality isn't optional—it's essential for AI readiness. As OpenAI engineer James Betker emphasizes, *"model behavior is determined by your dataset, nothing else"* (3 - non_int). Without high-quality data, generative AI initiatives will fail before they start.

Achieving great data quality requires collaboration between data, product, and engineering teams across products and user touchpoints. At Avo, we've seen how misaligned ownership and execution lead to wasted work, frustration, and lost revenue—making scalable governance more critical than ever.

The Data Governance Dilemma

In conversations with hundreds of data teams, a recurring challenge emerges: governing data at scale is unsustainable without the right approach. Central data teams become bottlenecks, struggling to maintain speed without oversight, so product teams bypass governance in order to move faster.

Even mature teams face this trade-off—sacrificing governance for speed when delivery deadlines take priority. This creates what we call the "data governance dilemma".

> The data governance dilemma is when teams are forced to choose between data quality and team velocity.

Teams either:
- Move fast and sacrifice data quality
- Enforce governance and slow everything down

Governance often loses out as product teams bypass cumbersome processes, reinforcing a vicious cycle of poor data quality. In response, data teams impose stricter controls, making governance even more rigid—only for teams to bypass it again.

Instead of driving strategic value, data teams are stuck firefighting data issues, wasting time that could be spent on data activation, business growth, AI, and personalization.

Breaking Free From the Cycle

To escape this dilemma, organizations need a scalable, self-serve governance model—one that ensures data quality without slowing teams down. Proven frameworks like Data Contracts and Data Mesh provide a foundation for this shift. However, while we've gotten behind data mesh, we've lacked a tactical guide to apply the data mesh principles in practice to event based data.

Organizations that implement scalable governance models see measurable improvements. For example, one financial institution reduced risk, accelerated innovation, and increased time-to-market for data solutions by 40% by implementing an end-to-end governance approach (4-McKinsey).

> Effective governance isn't about adding red tape — it's a catalyst for speed, agility, and business growth.

At Avo, we believe organizations shouldn't have to choose between bad data delivered quickly and good data slowed down by governance. There's a third option—one that empowers domain owners while ensuring central governance remains intact.

This white paper introduces a step-by-step recipe to implement self-serve governance, moving from chaotic data collection to structured, scalable governance. By applying the key ingredients and secret sauce outlined, organizations can achieve high-quality data at scale without sacrificing speed.

The Enablers of Data Governance at Scale

Data Contracts and Data Mesh are widely recognized among data practitioners. When applied to event-based data, together they can enable data governance at scale. Let's explore how each of them contributes to solving the data governance dilemma.

Data Contracts

The concept of data contracts originates from software engineering, where API contracts define clear agreements between services and has been adapted to data.

> Data contracts are formal agreements established between data producers and consumers to ensure quality, structure, and accountability (5 -Datacamp).

While early discussions of data contracts in governance emerged in the late 2010s, their adoption has surged as organizations seek scalable solutions to data quality challenges (6 - Dataliftoff). In practice, a data contract codifies expectations around a dataset's schema, semantic meaning, refresh frequency, quality metrics, and ownership.

Data Contracts in Event-Based Analytics

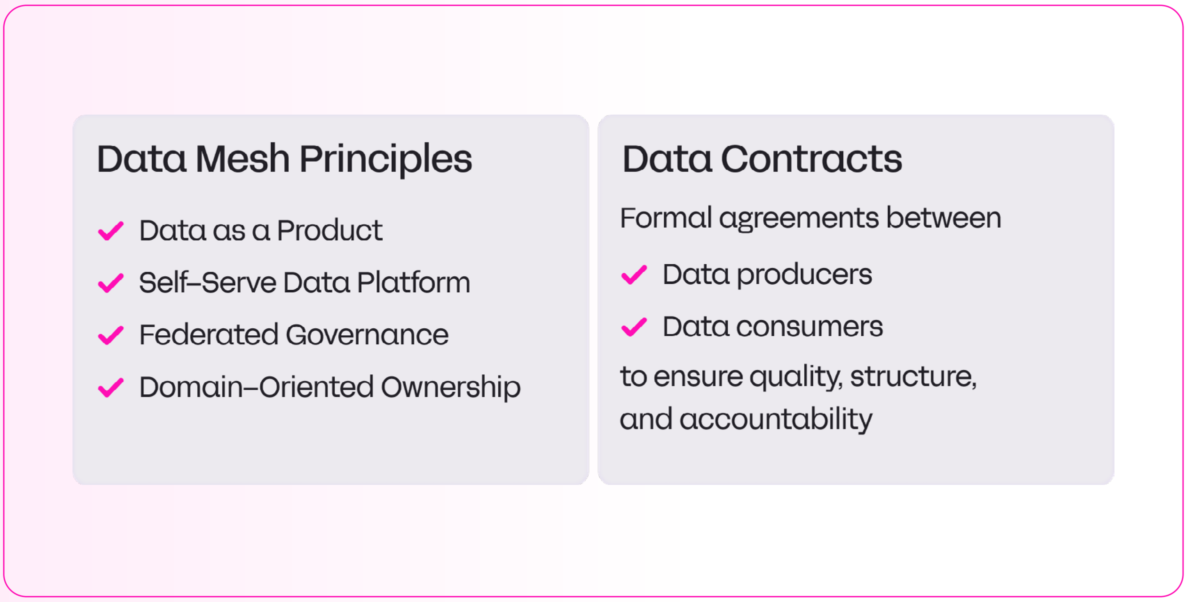
For event data, data contracts serve as a tool for collaboration, defining how teams generate and use data by answering key questions:

- **Ownership** – Who is responsible for the data?
- **Dependencies** – What systems or people depend on it?
- **Structure** – How should the data be structured, what kind of schema should it follow?
- **Lifecycle** – How, when, and where is the data created, processed, and consumed?
- **Impact** – What business metrics rely on it?

The Role of Data Contracts in Scalable Governance

Data contracts provide a structured, scalable way to improve governance, ensuring reliable, high-quality data without creating bottlenecks. By aligning expectations and assigning responsibility, they help streamline ownership and reduce friction.

Could they solve the data governance dilemma? Yes—but only with the right framework to support them.



Summary of Data Mesh principles and Data contracts

Data Mesh

The Data Mesh framework, first introduced by Zhamak Dehghani in 2019, has become a widely accepted approach for scaling data governance and analytics.

> Data mesh shifts data ownership to domain teams, enabling them to manage data as a product, while ensuring interoperability and federated governance at scale so organizations can build self-serve data cultures.

At its core, Data Mesh is built on four foundational principles (7 - Dehghani):

1. **Domain Ownership**
2. **Data as a Product**
3. **Self-Serve Data Platform**
4. **Federated Computational Governance**

Below, we'll explore each principle in detail, with a focus on how they apply to event-based analytics.

1. Domain Ownership: Bringing Data Responsibility to the Source

Traditional Challenge: A domain (8 - dbt) represents a specific area within a product or organization. Attached to a domain is a team —whether it's a product team with engineers and analysts or a finance team reporting to the CFO—that usually operates with some level of autonomy.

Despite this autonomy, most domain teams remain passive data consumers, relying on centralized data teams for insights. This creates bottlenecks, misalignment, and delays, as central teams struggle to collect, interpret and serve data for multiple domains.

The Data Mesh Approach: Instead of treating domain teams as data consumers, data mesh enables them to own and manage their data. This shift, known as "domain-driven ownership", ensures that teams who generate data are responsible for its accuracy, structure, and usability.

Event Data Applications: Looking at event data specifically, a product domain such as **"Search"** is responsible for defining, implementing, and analyzing event structures related to the Search experience of that product. By owning this process, domain teams ensure that data remains relevant and accessible, while reducing reliance on a central data team. Additionally, this ownership enhances data quality, as the teams generating the data are directly accountable for its governance and accuracy.

2. Data as a Product

Traditional Challenge: In conventional data architectures, data is often inaccessible, poorly documented, and difficult to trust. Gartner calls this "dark data"—information that is collected but rarely used (9 - Gartner).

The Data Mesh Approach: Data should be treated as a product and data consumers as customers. Great products succeed when they balance usability, desirability, feasibility, and business viability (10 - Cagan). To achieve this, organizations must establish clear ownership and objective measures to ensure data is consistently reliable and valuable. This includes defining functions like domain data product management, responsible for enforcing data standards and ensuring that data is delivered as a high-quality, usable product.

Event Data Applications: Event data should be designed with its end users and tools in mind, whether they are product managers or analysts in product analytics tools, or data practitioners building machine learning models on top of a data warehouse. Like a well-built product, data should be discoverable, trustworthy, and easy to use.

Ensuring alignment (11 - Avo) between data producers (developers) and data consumers (PMs, data practitioners) allows organizations to prioritize, structure, and maintain high-quality data effectively. This alignment also enhances data literacy, as deeper insight into how data is created and managed empowers teams to use it more effectively for decision-making.

3. Self-Serve Data Platform

Traditional Challenge: Data platforms are often centralized and complex, requiring teams to depend on data engineers for access, transformation, and provisioning. This slows down innovation and decision-making.

The Data Mesh Approach: A self-serve data platform provides domain teams with the infrastructure to manage their data autonomously — without relying on a central data engineering team.

Event Data Applications: For event data, non-technical data consumers should be able to easily access the data that they need to understand the experience of users interacting with their product. But having access to the data tool alone does not ensure self-serve data culture. In order to be self-sufficient, data consumers also need to be able to contribute to data collection, including defining the data structures. That will empower them to increase the insights accessible to themselves.

4. Federated Computational Governance

Traditional Challenge: Organizations often struggle to balance autonomy and consistency. Without governance, teams use inconsistent data structures; with too much centralized governance, teams are slowed down.

The Data Mesh Approach: Federated computational governance ensures that domain teams follow a shared, centrally defined framework, while still maintaining local decision-making power. This allows for scalability without losing control.

Event Data Applications: Event data governance should be seamlessly embedded into systems rather than enforced manually, allowing organizations to maintain consistency without adding friction to workflows. Contrary to centralized governance, the role of the federated governance team is not to ensure data quality, but to define how to model what constitutes quality and establish a framework for data quality that the domain teams can run with on their own. The federated computational governance should ensure event schemas, naming conventions, and data standards are consistently followed. By leveraging automated validation, schema registries, and stakeholder alignment, organizations can prevent data drift and uphold high-quality analytics without disrupting development processes.

Stages of Data Culture Maturity

Implementing data contracts and data mesh provides the foundation for scalable data cultures, with governance that enables data literacy and data quality, rather than being a burden.

Scaling a data culture doesn't happen overnight. At Avo, with decades of hands-on experience and insights from thousands of data teams, we've found that organizations—whether entire companies or individual teams—typically progress through the following stages of data culture maturity:

1. Winging it
2. The Wild West
3. Centralized Governance
4. Self-Serve Governance

The stages of data culture maturity

While these stages describe the evolution of data culture as a whole, the challenges and processes we outline below focus specifically on event data collection. Understanding its progression within each stage illustrates the path from data chaos to scalable governance, helping organizations assess their current state and plan their next steps.

Stage 1 – Winging it: No Data

Teams operate without structured data collection, relying on qualitative insights, experience and intuition rather than analytics.

Processes: None

No formal data collection and hence no governance.

Challenges: Flying Blind

Decision-making is subjective and inconsistent; data is not repeatable or reliable.

Transition Trigger

As the organization and user base grows, as does the appetite for quantitative data. As products grow in complexity, the need for a bigger-picture understanding of user behaviour and the impact of strategic decisions increases. Leadership demands data-driven insights, pushing teams toward more structured data collection.

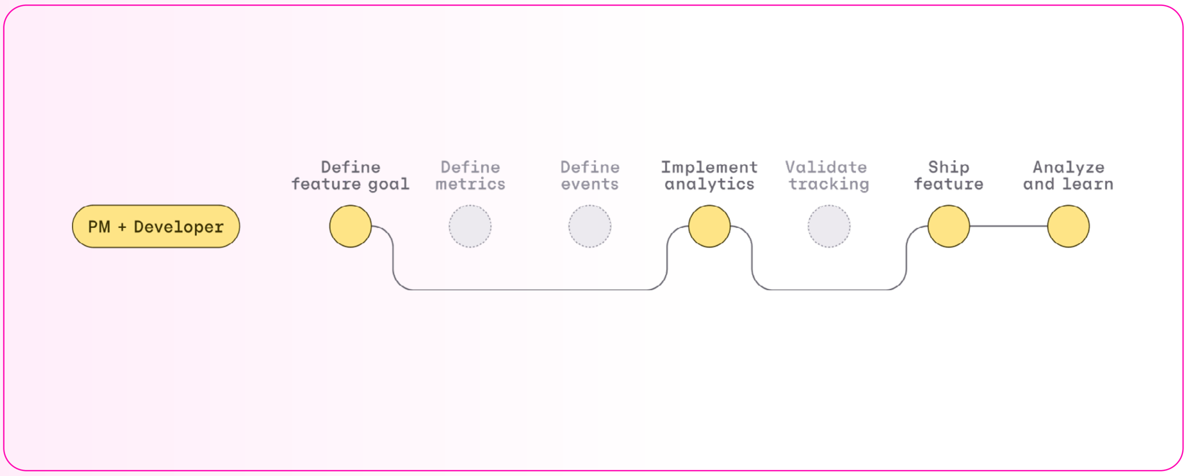
Stage 2 – The Wild West: Bad Data

The Wild West data culture maturity stage

Teams collect data to satisfy their own short-term need for analysis and insights of isolated features. Data is defined and collected on an as-needed, per-team basis.

Processes: Informal

Product teams ship event collection independently with no standardization of how data should be collected across teams and product areas. There are little formal processes or documentation for defining event taxonomies. Quality control of event implementation happens if and only if the contributor has the data maturity, skill and motivation.



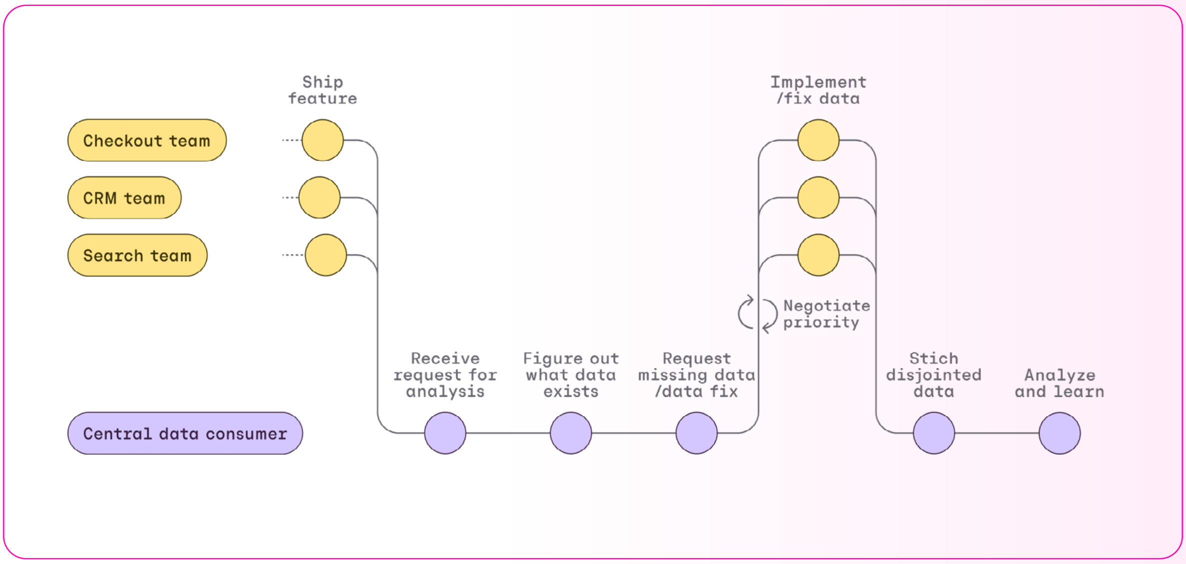
Informal Wild West workflow, where important steps are skipped

Challenges: Inability to Build Org-wide Data Products

While the Wild West workflow may work for a single feature within one team, the chaos becomes painfully clear once the need arises for leveraging data from multiple teams. In the absence of a formalized data collection process, consolidating data for data products that leverage data across teams requires tremendous overhead and manual effort.

The data practitioner tasked with creating a central data product must first determine what data exists, often finding inconsistencies, gaps, or duplication. The lack of standardization means different teams track similar user actions in conflicting ways, making it labor-intensive or impossible to stitch together a cohesive dataset.

Fixing these issues requires data practitioners to chase down product teams for clarifications and corrections. But these requests compete with ongoing development priorities, leading to delays, frustration, and mis-aligned expectations.



Reactive damage control: the consequences of poor alignment

A single metric might be tracked with seven different event names, forcing complex SQL workarounds. This “Frankenstein SQL query” doesn’t just persist — it grows. With each new request, the data team patches it again, adjusting for the latest data inconsistencies instead of solving the root problem.

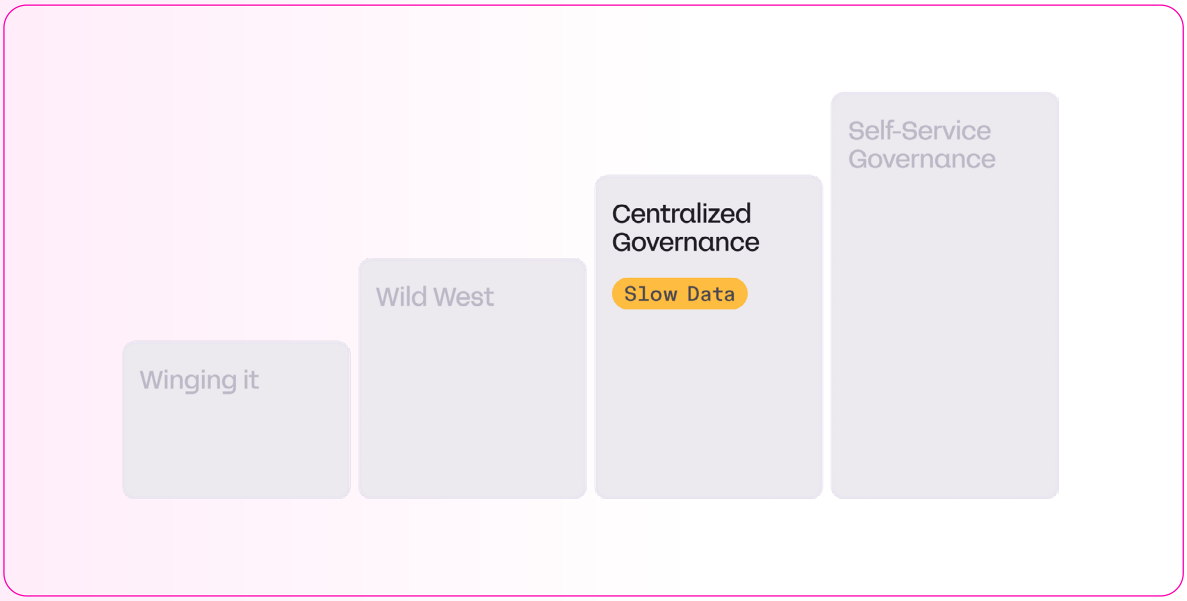
> Instead of driving insights, teams in the Wild West are stuck in reactive damage control, trying to fix broken data rather than leveraging it for innovation.

Transition Trigger

As organisations shift to actively seeking scalable data products, data quality emerges as a critical blocker. At this stage, the need for consistent data becomes undeniable, and the inefficiencies of reactively patching up bad data become unacceptable. This leads to organizational buy-in for top-down governance initiatives driven by data teams.

Stage 3 – Centralized Governance: Slow Data

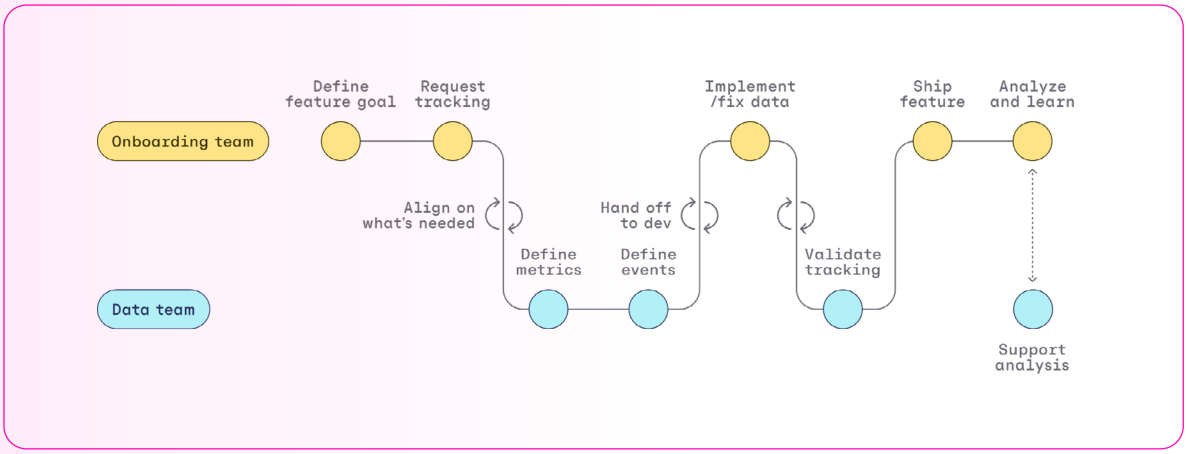
A central data team introduces governance and enforces it centrally to ensure that data is collected consistently across the organization.



The Centralized Governance data culture maturity stage

Processes: Centralized

In a centralized governance model, organizations introduce formal processes for the release of new data structures. Product teams now request tracking from the central data team – which in turn translates the feature goal into metrics and data structures and ensures developers implement standardized tracking.

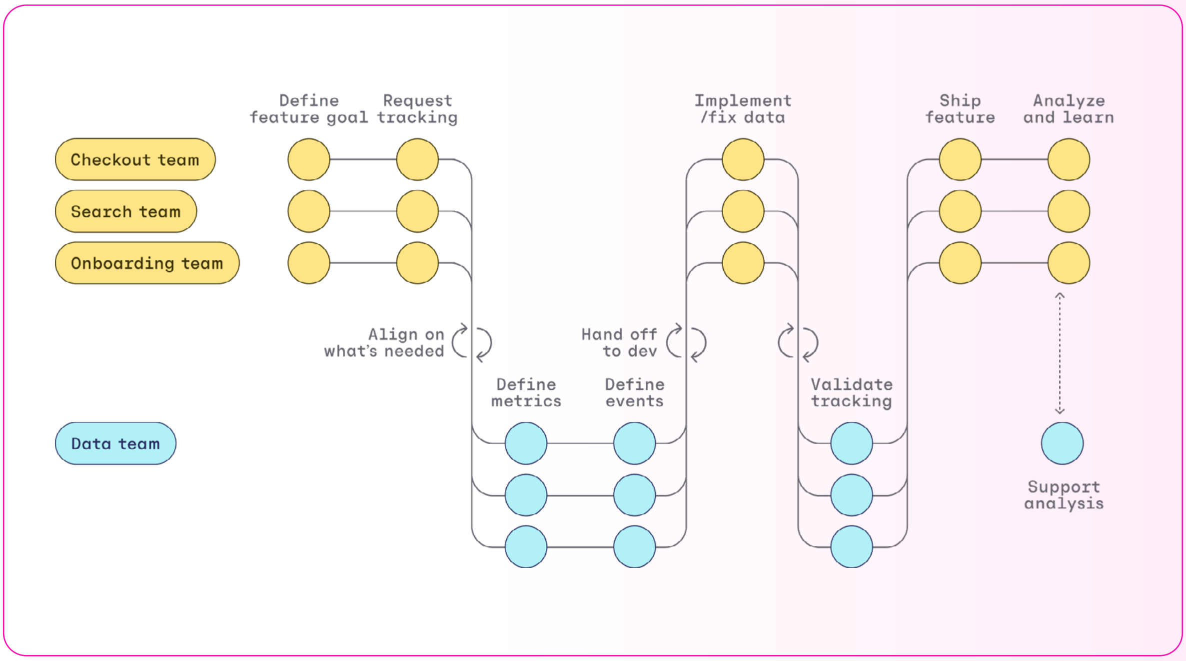


Centralized workflow, where the data team handles all data definitions

This approach improves data consistency and quality but introduces new challenges around agility.

Challenges: Central Data Team Becomes a Bottleneck

In the Centralized Governance stage, all new data collection specs should be defined by the central data team. This ensures high data quality — but at the cost of speed. The process of reviewing and approving data collection requests becomes slow and tedious, often taking weeks instead of days. Data teams find themselves spending more time in alignment meetings and context-switching rather than enabling teams to move forward efficiently.



Data team bottleneck: the consequences of centrally defined data

Transition Trigger

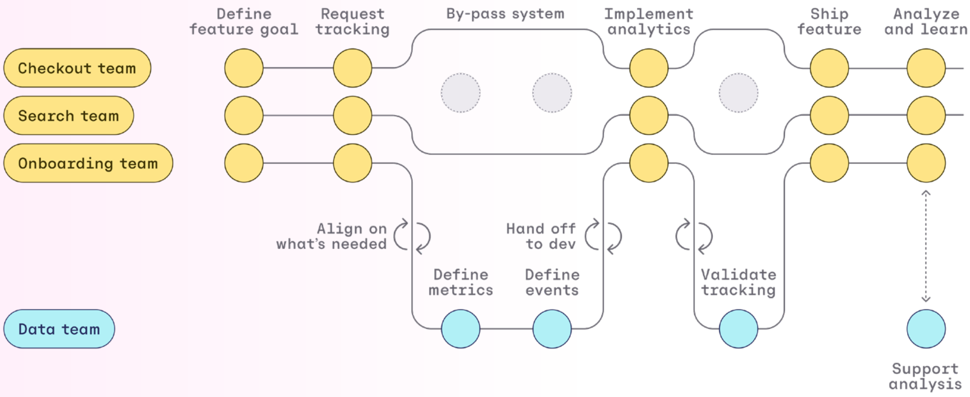
The friction between autonomy and governance pushes teams to explore self-serve solutions to retain the autonomy of the Wild West approach, without compromising the quality of centralized governance.

Data Governance Dilemma: Centrally Governing the Wild West

While most companies at scale institute data governance, few of them manage to fully rein in all teams to adhere to the centralized governance model. This brings us back to the data governance dilemma—where teams must choose between moving fast and risk broken data or enforcing governance that slows everything down.

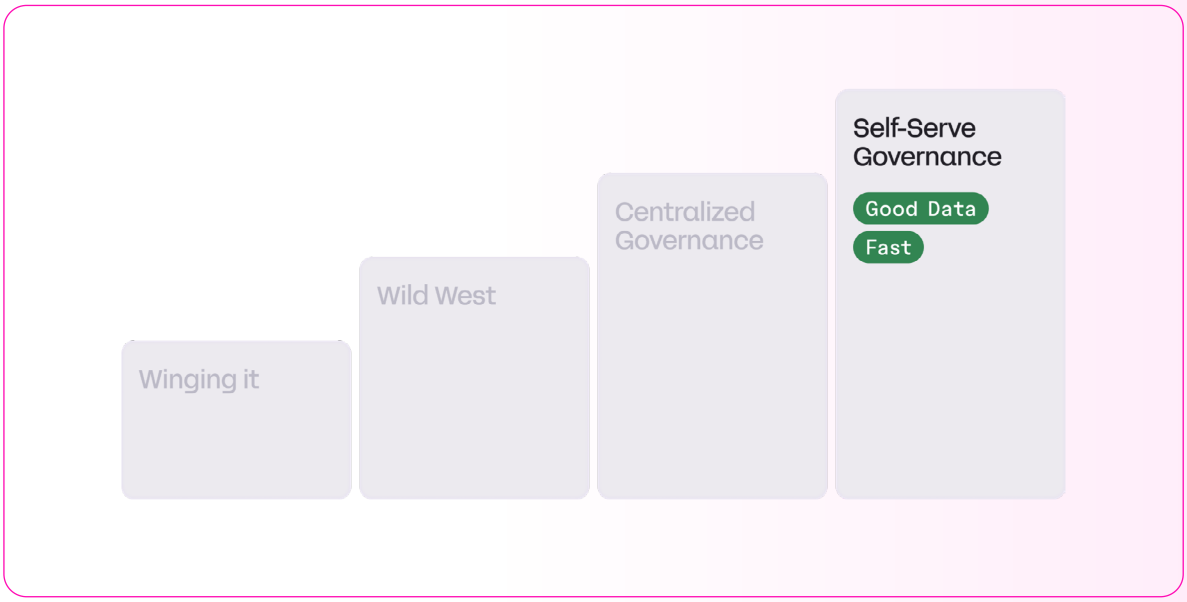
Due to the inefficient nature of centralized governance, most data-driven organisations find themselves stuck between the wild west and centralized governance. Some teams or departments collaborate with the central team while others bypass the system to avoid the bottleneck—leading to shadow IT (12 - IBM).

This leaves even those data teams with dedicated teams for governance stuck in reactive damage-control when trying to ensure consistent data across their organization.



Centrally governing the Wild West, with systems in place that some teams bypass

Stage 4 – Self-Serve Governance: Good Data, Fast



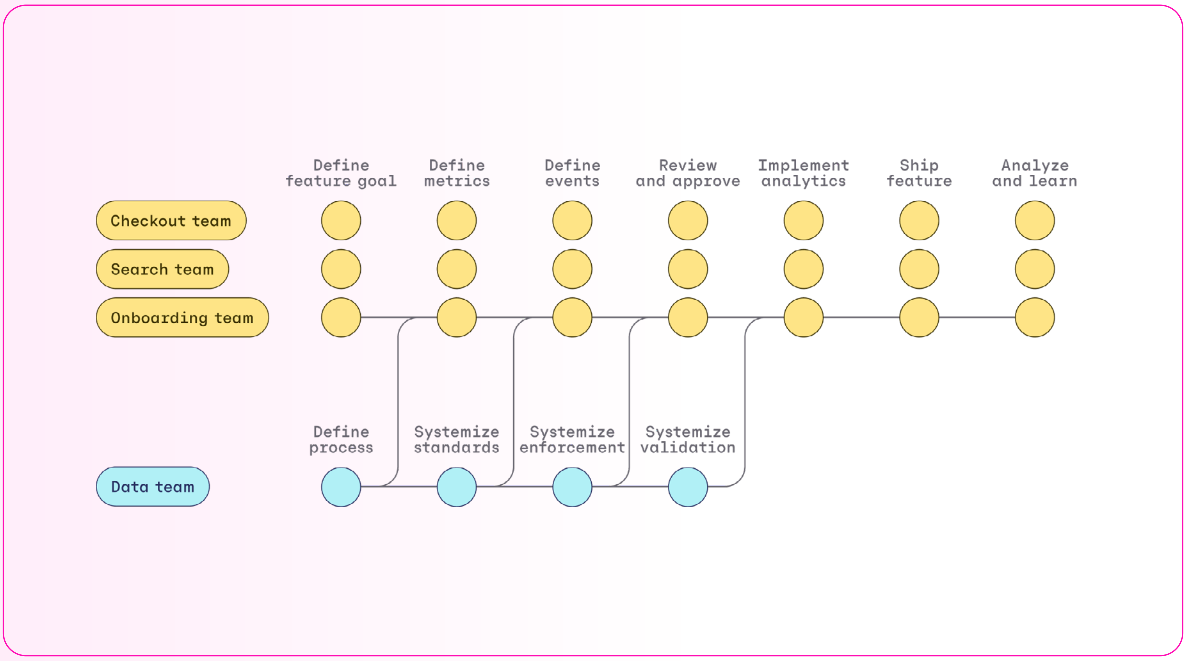
Teams manage their own data workflows while governance is enforced through systems and processes, instituted by the central governance team.

This is done through the application of data mesh principles and data contracts or other structured working agreements.

Processes: Centrally Defined - Locally Enforced

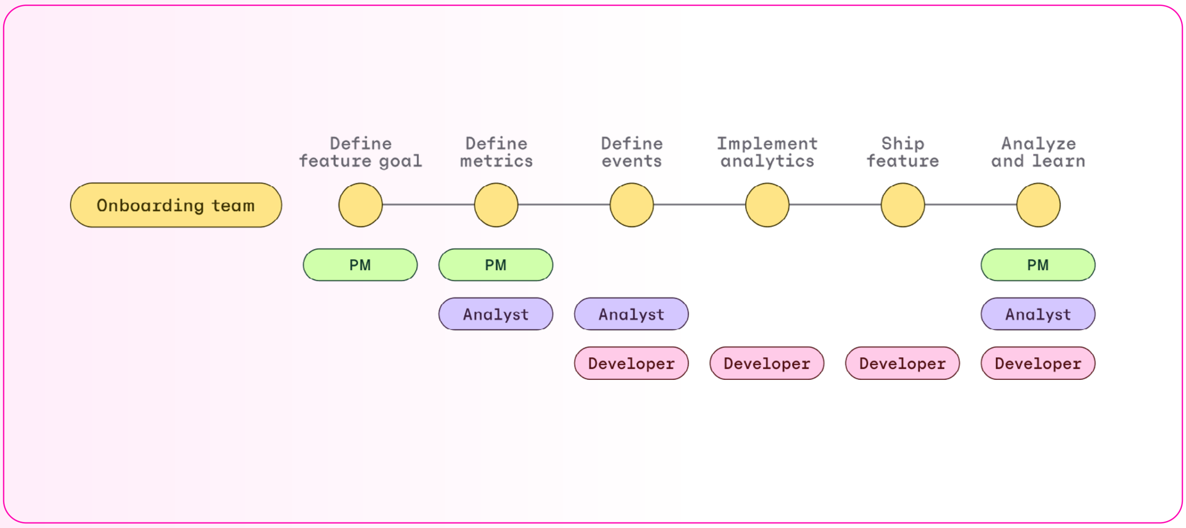
The federated governance function, rather than manually taking on the work of defining event structures and validating them, establishes systems and processes that enable individual teams to define and ship data that's consistent across the organization.

Teams manage their own data workflows while governance is enforced through automated guardrails and peer reviews. Each role contributes at different stages, ensuring collaboration, accountability, and consistency across the process.



Self-serve governance workflow: where the data team provides systems for domains

Achieving velocity in data collection without sacrificing data quality directly depends on applying data mesh principles and reinforcing them with data contracts or structured agreements.



Role participation in the end-to-end workflow owned by domains

Data Mesh Principles

- ✓ Data as a Product
- ✓ Self–Serve Data Platform
- ✓ Federated Governance
- ✓ Domain–Oriented Ownership

Data Contracts

Formal agreements between

- ✓ Data producers
- ✓ Data consumers

to ensure quality, structure, and accountability

Summary of Data Mesh principles and Data contracts

Benefits

When executed effectively, the self-serve governance model shifts from being a challenge—like previous stages—to delivering only benefits. The key advantages include:

- **Speed:** Teams ship analytics faster without sacrificing data quality, thanks to structured workflows and guardrails.
- **Empowerment:** Teams feel more ownership and confidence in data when they actively participate in its creation
- **Scalability:** Organizations can leverage high-quality, consistent data across teams, enabling faster insights and greater impact.
- **Impact:** With less need for manual quality control, the central data team can focus on proactive, high-impact initiatives instead of reactive fixes.

Recipe for Data Mesh for Event Based Data

To overcome centralized governance bottlenecks and build a scalable, self-serve data culture, organizations must shift from manual enforcement to automated, systematic processes that ensure good data by default. To get there, consider the Data Mesh Principles and Data Contracts, important frameworks to benchmark with while building a data product management function. This recipe is a tactical guide to actually achieve a culture where you have Data Mesh and Data Contracts in place in practice, not just as theoretical concepts.

The key ingredients are an important foundation for successful data collection at any scale, but to really reach data reliability at scale there's a secret sauce—the advanced techniques that make it scalable, future-proof, and adaptable.

Organizations that level up and eventually automate their governance workflows towards reaching Data Mesh and Data Contracts ensure efficiency, flexibility, and long-term resilience in self-serve data management.

The Key Ingredients		The Secret Sauce
☑ Design data based on use cases	--->	☑ Document metrics with event definitions
☑ Standardize naming and abstraction levels	--->	☑ Enforce data standards
☑ Single source of truth	--->	☑ Source control and sync schemas
☑ Establish & document ownership	--->	☑ Accountability systems and alerts
☑ Data observability and alerting	--->	☑ Prevent issues upstream

The recipe for data mesh for event based data

The key ingredients

The foundations to set your data org up for success

☐ Design data based on use cases

Action
Shift from data design based on coverage to a goal-driven approach. Instead of asking, "Which events do we need?", ask, "What is the goal, and how will we measure success?"

Unlocks
Ensures teams collect only the most relevant data, prioritize effectively, and give engineers clear purpose and ownership over event data decisions.

☐ Standardize naming and abstraction levels

Action
Define standard naming frameworks (casing, tense, word order and vocabulary). Name and structure events based on user actions; deep or shallow depending on use case (13 - Avo).

Unlocks
Ensures consistency across products, platforms, events and code paths. Ensures human readable events while eliminating duplicate and inconsistent events.

☐ Single source of truth

Action
Centralize event definitions in one authoritative source instead of scattered across spreadsheets, Slack, or Jira tickets.

Unlocks
Ensures teams reuse existing event definitions instead of reinventing them. Enables alignment, and a reliable reference for how data should be structured.

☐ Establish & document ownership

Action
Clearly define who owns each event's definition and implementation, as well as who will be impacted by changes.

Unlocks
Ensures awareness that changing data could impact stakeholders, preventing disruptions to downstream teams

☐ Data observability and alerting

Action
Implement automated observability to detect deviations from expected data in real-time.

Unlocks
Ensures fast issue detection and resolution, preserving data reliability and preventing faulty decisions.

The secret sauce

Advanced techniques to apply as you scale

☐ Document metrics with event definitions

Action
Maintain a metrics store that not only documents event schemas but also tracks how each event is used in reporting and analysis

Unlocks
Creates a single source of truth for not only event schemas but also metric definitions, ensuring consistency across all teams.

☐ Enforce data standards

Action
Automatically validate whether event specs fulfill event naming framework.

Unlocks
Allows anyone to design data according to standards. Reduces redundant reviews, allowing data practitioners to focus on higher-value work.

☐ Source control and sync schemas

Action
Implement a schema documentation solution that enables parallel collaboration and automatic syncs to downstream schemas.

Unlocks
Prevents bottlenecks and ensures clarity and consistency in downstream applications. Domains can collaborate on their data like their code.

☐ Accountability systems and alerts

Action
Configure automated alerts to notify event owners and stakeholders when their data is modified or at risk of breaking.

Unlocks
Prevents unexpected changes that could disrupt dashboards or critical business functions, ensuring data integrity at all times.

☐ Prevent issues upstream

Action
Equip engineers with direct translations from data schemas to code and tools for validation.

Unlocks
Ensures data structures are implemented exactly as designed so that issues are not only caught, but prevented.

Good Data Product Management / Bad Data Product Management

With this recipe in mind, let's outline key benchmarks of good vs. bad data product management, highlighting the symptoms of missing foundations versus successful implementation:

Bad data product management tracks everything and measures nothing. Teams collect generic events like *"button clicked"* without context, drowning in noise. Data feels abundant but isn't useful.

Good data product management tracks what matters. Events align with key user journeys, ensuring data is actionable, insightful, and meaningful. Engineers know why they track events, and analysts trust the insights.

Bad data product management documents events but not how they contribute to metrics. No one knows what an *"active user"* actually means. Every analysis starts with debating definitions instead of extracting insights.

Good data product management connects events to business outcomes. Metrics are clearly defined, structured, and machine-readable. Teams know not just what happened, but why it matters.

Bad data product management is messy.
The same event appears under multiple names *("account created" vs. "account created - email")*, creating duplicates and confusion. Teams waste time cleaning up data instead of using it.

Good data product management enforces consistent naming and level of abstraction (when an event should be one event vs. many, with a property to distinguish between the cases). Events are structured, intentional, and easy to reuse, keeping data clean and scalable.

Bad data product management lives in spreadsheets, Jira tickets and Slack messages.
Schemas are manually copied across tools, creating misalignment and outdated versions. No one knows which version is correct.

Good data product management ensures schema changes sync automatically across all registries. Everyone works from one reliable source of truth, reducing misalignment and manual effort.

Bad data product management allows changes without stakeholder alignment. One team removes a field, breaking a dashboard for another team's dashboards, discovered two months later. No one was notified.

Good data product management makes ownership explicit. Stakeholders are automatically notified when schema changes impact them, preventing disruptions before they happen.

Bad data product management reacts too late.
Data collection is broken for months before anyone notices. Business decisions are made on bad data.

Good data product management catches issues before they impact decisions. Data quality tools detect errors in development (not when they've hit production), ensuring data quality at scale.

In summary:
Bad data product management creates chaos—collecting everything, understanding nothing. Good data product management scales effortlessly—structured, people-first, automated, and built for trust and data literacy.

Implementing the Recipe for a Better Data Product Management

The key ingredients and secret sauce lay the foundation for self-serve governance—but how do we implement them effectively? How do we choose the right processes and tools to maximize their impact?

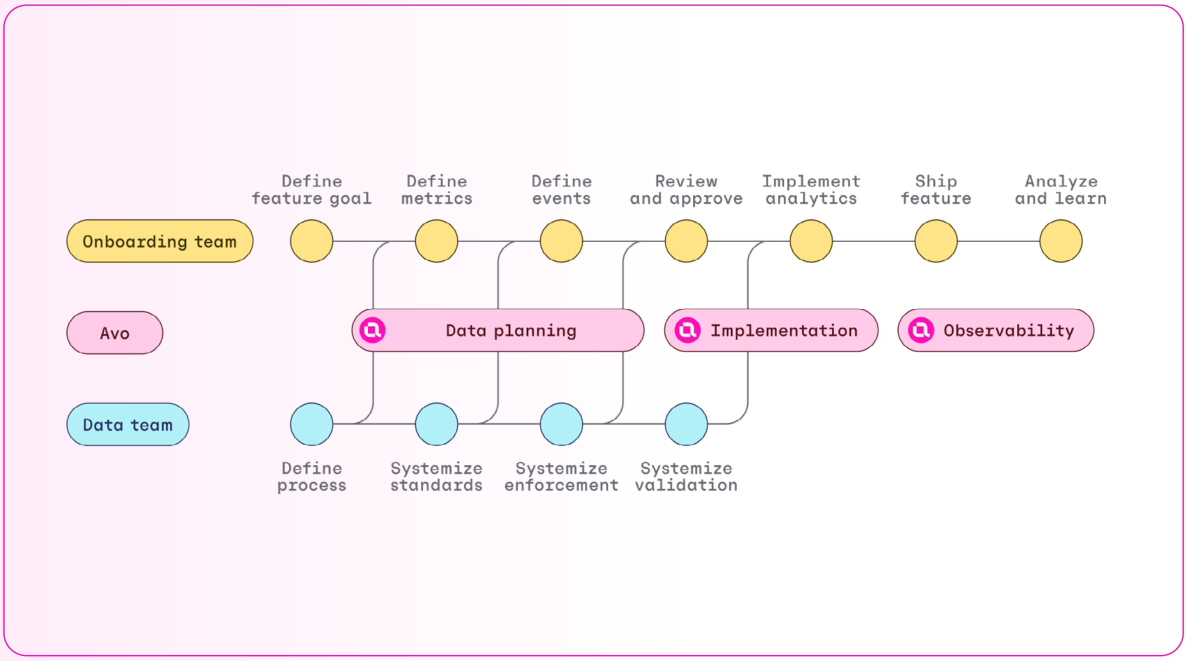
> Building a scalable governance model begins with cultural alignment. The organization must establish a shared vision for data quality across leadership, disciplines, and domains, and key roles—including data practitioners, product managers, and engineers.

Transitioning from the Wild West to self-serve governance requires more than just tools and automation — it hinges on organizational buy-in.

The principles outlined in this framework have been developed through extensive collaboration with thousands of data practitioners, product engineers, and product managers who have successfully implemented these methodologies. Many organizations have relied on custom-built solutions to manage governance—including the team behind Avo, who built internal tools at every company they worked at before Avo. While these solutions offer short-term fixes, they lack scalability, automation, and, most importantly long-term maintainability. As priorities shift and employees move on, homegrown tools rarely evolve to meet the demands of a growing organization.

Avo's Role in Scaling Data Quality

Avo enables Enterprises to achieve event data quality at scale by moving data quality management upstream. The Avo platform facilitates cross-functional collaboration, automates enforcement of standards and ownership, and ensures scalable, efficient data governance.



Avo supporting the self-serve governance workflow

Key capabilities include:

1. Schema Change Management

Avo replaces static spreadsheets with source-controlled schema management. In Avo, teams govern their event schemas in a structured workflow, with built-in guardrails, assigned reviews and automated audits. By making event schemas accessible and managing changes in a central, collaborative platform, teams ensure changes are standardized, reviewed and aligned across teams before implementation.

Results:

- Consistent data structures across teams and user touchpoints
- Reduced overhead in data collection planning and review
- Faster onboarding for new team members
- Higher confidence in data

2. Implementation & Validation

Avo enables data and engineering teams to align on event definitions, and provides tools for fast, error-proof implementation. The Avo platform serves as a single source of truth for event structures, generating clear data collection specs and providing validation tools that developers can integrate into their codebase, from schema payload observability to code generation. By aligning data and engineering teams and systemizing implementation and validation, teams ensure implemented data collection matches the defined schema.

Results:

- Fewer data collection errors and implementation bottlenecks
- Reduced developer time spent on QA and troubleshooting
- Consistent, reliable data without back-and-forth adjustments

3. Data Observability & Alerting

Avo Inspector monitors event data collection across development and production, detecting schema violations in real time, including unexpected nulls and type mismatches. Avo provides proactive alerts, enabling teams to catch issues early—before they impact business-critical data products and downstream systems.

Results:

- Faster detection and resolution of data inconsistencies
- Reduced time spent troubleshooting and patching broken data
- Higher data reliability for decision-making

Case Studies: Reaching Data Quality at Scale with Avo

How Wolt Scaled from Data Chaos to a Unified Data Strategy

> ”Getting data foundations right is critical. Before Avo, data collection was fragmented, and aligning data across teams was a massive challenge. Now, we have a structured system that ensures accuracy and speed in decision-making.”
>
> — Jacopo Himberg,
> Director of Data @ Wolt (DoorDash subsidiary)

Industry: Food Delivery
Company Size: 10,000+
Tech Stack: Snowflake, BigQuery, Looker, Tableau, Mixpanel, Avo

Challenge

Wolt, a rapidly growing food delivery platform, grew from 25 employees in 2016 to over 10,000+ in 2024, with dynamic product teams relying on data-driven experimentation and decision-making.

With expansion across 25+ countries, its data operations struggled with:

- **Decentralized and Siloed Data Collection** – Engineers and analysts lacked a unified approach to define and collect event data. Even with the introduction of event definitions in YAML files, it was difficult to maintain consistent schemas across teams.
- **Slow Implementation** – Event structure changes required manual coordination across multiple teams.
- **Unreliable Data** – Engineers built data collection frameworks but had no way to verify implementation accuracy. Data issues were often discovered months after experiments had run, leading to wasted resources

- **Overwhelming Complexity** – Managing 800+ app events and 230+ databases led to data inconsistencies.

Solution

Wolt implemented Avo to transform its approach to event data collection and data governance:

- **Single Source of Truth** – Avo centralized data collection specs, eliminating confusion.
- **Developer-first Data Collection** – Engineers integrated Avo with their workflow, ensuring accurate event implementation.
- **Automated Validation** – Teams could now verify with Avo Inspector and Avo Codegen if event structures were implemented correctly before launching experiments.
- **Scalability Beyond Apps** – Avo helped Wolt transition towards **data contracts** for both event analytics and warehouse governance. A scale that YAML files on git were not able to provide.

Results

By embedding data contracts into engineering workflows, Wolt improved data accuracy, developer experience, and speed to insights:

- **90% Faster Event Structure Implementation** – From weeks of coordination to seamless integration.
- **80% Reduction in Data Quality Issues** – Automated validation ensured data integrity before release.
- **Increased Analytics Agility** – Faster access to reliable data for decision-making and experimentation.
- **Scalable Data Governance** – Avo helped Wolt move towards **codified data contracts** across 230+ databases.

With Avo, Wolt transformed its data collection from fragmented YAML files to a fully governed, developer-friendly system, enabling faster product iteration, experimentation, and strategic decision-making. Read the full case study on avo.app.

How Delivery Hero Went from Schema Management Chaos to Alignment in a Single Source of Truth

> "Avo is the true foundation for our data solutions. Before Avo, aligning data collection across all brands took months. Now, it takes about a week."
>
> — Cathy Wong
> Data Governance Manager @ Delivery Hero

Industry: Food Delivery
Company Size: 10,000+
Tech Stack: Amplitude, BigQuery, Looker, Tableau, Avo

Challenge

Delivery Hero operates 12+ independent brands across 70+ countries, each with its own data schemas.

The central Perseus data platform team needed to maintain governance at scale, but faced:

- **Decentralized Data Collection** – Each brand maintained separate tracking documentation, leading to misalignment
- **Slow Implementation** – New events required lengthy approvals across multiple teams
- **Data Silos** – Local teams lacked visibility into how other brands structured events
- **Time-consuming Debugging** – Analysts struggled to trace errors across inconsistent schemas

Solution

Delivery Hero adopted **Avo** to:

✓ **Unify Schemas** under a single source of truth

- **Enable Federated Governance** with flexibility for local teams
- **Speed Up Event Implementation** without back-and-forth delays
- **Reduce Errors** by improving visibility across teams

With Avo, Delivery Hero moved from decentralized Google Docs chaos to scalable, structured governance.

Results

The results, in addition to improved data quality, were significant time savings and reduced back and forth:

- **91% Faster Data Collection Process** – From months to a week
- **80% Faster Debugging** – Issues resolved in 1–2 days
- **Increased Alignment** – Teams avoid redundant work and conflicting schemas

Full case study available soon on avo.app.

Conclusion

Organizations no longer need to be stuck in the data governance dilemma and choose between fast but broken data or slow but high-quality data. With the right processes, automation, and governance models, companies can scale self-serve data governance—ensuring data remains trustworthy, accessible, and aligned across teams.

By implementing Data Mesh principles and Data Contracts, organizations shift from reactive damage control to proactive data quality management. The result? Faster decision-making, reduced data inconsistencies, and the ability to leverage data for AI, personalization, and business growth.

Companies like Wolt, Delivery Hero, and Moody's have demonstrated the impact of scalable data governance—accelerating implementation, reducing errors, streamlining collaboration across teams, and, most importantly, maximizing data's impact on business success.

Ready to move from theory to execution to build data quality at scale?

Whether building internal solutions or leveraging tools like Avo, organizations must focus on automating governance, aligning stakeholders, and embedding data quality into workflows—ensuring good data by default at every scale.

To explore how Avo can support your data governance strategy, contact us at hi@avo.app.

The recipe for data mesh for event based data

The Key Ingredients		The Secret Sauce
☐ Design data based on use cases	---->	☐ Document metrics with event definitions
☐ Standardize naming and abstraction levels	---->	☐ Enforce data standards
☐ Single source of truth	---->	☐ Source control and sync schemas
☐ Establish & document ownership	---->	☐ Accountability systems and alerts
☐ Data observability and alerting	---->	☐ Prevent issues upstream

About the Authors

Stefania Olafsdottir – CEO & Co-Founder @ Avo
Klara Arnalds – Customer Experience Architect @ Avo
Thora Gudfinnsdottir – Principal Product Manager @ Avo

Collectively, Stefania, Klara and Thora bring decades of experience in managing data workflows and tools.

Stefania is a philosopher and mathematician, turned genetics researcher, turned data leader, turned founder. She has been a data practitioner for 15 years. She was the founding data person and became Head of Data Science and Growth Lead at QuizUp (100m users, backed by Sequoia and Tencent). This led her to start Avo to inspire better data cultures so teams could build better user experiences.

Thora has experienced the challenges of event data at scale firsthand since 2014, working as a data scientist at a fast-growing mobile company. Ever since, she's been passionate about data quality and managing tracking plans. Since joining Avo as its first employee, she's worn nearly every hat. Over the past few years, her focus has increasingly shifted toward building great user experience for data practitioners.

Klara joined Avo driven by her passion for solving complex user problems through systems thinking. Evolving from a product designer to her current role as Customer Experience Architect, she consistently applies a research-driven, user-centric approach to product development. She has worked with enterprises to bring design system principles to data design, helping them structure their processes for scalable and effective data governance.

Thank you to Glenn Vanderlinden, Maura Church, Fredric Lundgren, Claire Armstrong, Qiaoran Abbate, Giorgio Terreni, Cathy Wong, Timo Dechau, Thomas in't Veld, Jacopo Himberg, and Jake Winter for your contributions to the ideas in this paper.

Further Reading

- Olafsdottir, S. Building data teams. Tips from Erik Bernhardsson's journey at Spotify and beyond. Avo blog, 2022. Read the article
- Olafsdottir, S. The business impacts of data quality issues —what they look like and how to think about fixes. Mixpanel blog, 2025. Read the article
- How Wolt drove a data culture revolution with Avo and Mixpanel. Avo case study. Read the case study