Customer Data, Reporting & Insights Platform

A Full PM View: Strategy · Architecture · Roadmap · Execution

Farzad Rajabi | hc.farzad@gmail.com | April 2026

The Problem Worth Solving

Sinch generates massive volumes of communication data. Today, almost none of it reaches customers in a usable form.

12+

product silos, each with its own schema & API quirks

175K+

businesses whose data is fragmented across products

0

unified cross-product data APIs available today

What this costs customers today:

- Developers write 3-week ETL pipelines to build what the platform should provide
- Finance teams reconcile invoices manually against 4 different product dashboards
- Marketing ops build Friday spreadsheets because no single delivery rate view exists
- AI, alerts, and intelligence capabilities are impossible without a unified data layer first

The Platform Vision

What we are building — and what we are not

STRATEGY

Every business using Sinch has immediate, complete, and trustworthy access to their communication data — in whatever form they need it — so they can understand what's working, act on what isn't, and prove the value of every message they send.

HORIZON 1 · Months 1–12

See Everything

Any customer sees all Sinch data — across all products — in one place, accurately, without building anything themselves.

HORIZON 2 · Months 13–24

Act on Everything

Platform proactively surfaces what matters, integrates with tools customers already use, and enables action — not just observation.

HORIZON 3 · Months 25+

Understand Everything

The platform tells customers not just what happened, but why — and what to do about it. Powered by AI on a trusted data foundation.

Explicitly NOT: a CDP · a BI tool · a marketing automation platform · a real-time streaming infrastructure

Who We're Building For

Five distinct personas — validated through deep discovery research

DISCOVERY

Maya

PHASE 1 PRIORITY

Campaign Operations Manager

"Builds a Friday spreadsheet combining 3 product dashboards because no unified delivery view exists"

Arjun

PHASE 1 PRIORITY

Platform Developer

"Writes 3 weeks of ETL code to build what a single API endpoint should provide"

Stefan

PHASE 1 PARALLEL

Finance & Compliance Director

"Reconciles invoices manually across 4 dashboards — can't find the billing data he needs"

Priya

PHASE 2

BI / Data Engineer

"No native Sinch connector — rebuilds data pipelines from scratch for every product"

Leila

PHASE 2

VP Analytics / Executive

"Can't get a single cross-channel ROI number without commissioning a 2-week analysis"

DISCOVERY

The Journey We're Replacing

Before the platform — and what changes when it exists

Maya · Campaign Operations Manager

AS-IS · Every Monday morning

Opens 3 separate dashboards

Downloads CSVs from SMS, Email, and Conversation API UIs

Manually merges in Excel — 2–3 hours of work

Builds a slide before the Friday standup, hoping the numbers match

▼ WITH THE PLATFORM

TO-BE · Month 4 onwards

Single dashboard: SMS, Email, WhatsApp — one view, same definitions

Alert fires if delivery rate drops below threshold.

Scheduled report lands in stakeholder inboxes every Monday at 8 am

Friday number matches Monday number. No reconciliation.

Arjun · Data / Platform Engineer

AS-IS · Every integration cycle

Writes a custom ETL pipeline per Sinch product

Each product has different auth, schema, and pagination

Pipeline breaks silently when any product updates its API

Spends 3 weeks on what a single REST call should provide

▼ WITH THE PLATFORM

TO-BE · Month 3 onwards

One authenticated API call returns events across all channels

Canonical schema is stable and versioned — 180-day notice for breaking changes

Bulk export: async job → signed Parquet URL → SHA-256 checksum

First pipeline running in < 30 minutes from account provisioning

North Star Metric

What we optimize for — and why WDAA outperforms the alternatives

METRICS

Weekly Data-Active Accounts (WDAA)

Distinct Sinch customer accounts that perform ≥ 1 meaningful data action (query, export, dashboard view, API call, scheduled report delivery) in a rolling 7-day window

ACTIVATION

- Time-to-first-API-call ≤ 30 min
- Onboarding completion rate ≥ 70%
- First meaningful action < Day 7

ENGAGEMENT

- D30 retention ≥ 55%
- Actions per active account / week
- Scheduled report open rate ≥ 40%

EXPANSION

- ≥ 2 active users per account (30%)
- Multi-channel query rate ≥ 40%
- WDAA ⟶ 500 accounts by Month 12

Internal Health Metric: Products Actively Contributing Data (PACD) — target ≥ 6 products by Month 12

Competitive Landscape

Where competitors leave the market open — and where Sinch wins

COMPETITIVE ANALYSIS

Capability	Twilio	Infobip	Bird	Sinch CDP (this platform)
Unified cross-product API	✗ 3 products	✓ Unified	✗ Basic	✓ All channels
Data retention	120-day cap	Long-term	Limited	Configurable
Warehouse connectors	✗ None native	✗ None	✗ None	✓ Phase 2 (Fivetran)
Billing reconciliation	Manual	Partial	✗	✓ Native
AI / NLQ interface	✗	Roadmap	✗	✓ Phase 2 beta

Key gap No competitor combines cross-channel unified data + billing reconciliation + warehouse connectors + an AI query interface in a single platform. This could be the differentiation gap.

STRATEGIC ANALYSIS

Honest Self-Assessment: SWOT

What we have going for us — and what can derail us if we're not deliberate about it

STRENGTHS

- Data at scale — billions of events across 175K+ customers
- Deep carrier relationships = authoritative delivery data
- No cold start — 175K existing accounts to activate
- Multi-channel breadth: SMS, Email, Voice, WhatsApp in one portfolio

WEAKNESSES

- Platform doesn't exist yet — building from zero while competitors iterate
- Acquired products use incompatible event schemas (normalization effort)
- No cross-product PM authority — influence without mandate
- Data quality risk: bad events from product teams are invisible until customers notice

OPPORTUNITIES

- No competitor has solved cross-channel unified reporting — gap is real
- Enterprise data engineering is genuinely underserved across all CPaaS vendors
- Warehouse integration creates high switching costs once built
- Each new acquisition = new data source → platform becomes more valuable over time

THREATS

- Internal product teams may resist schema standardization (loss of autonomy)
- Twilio could invest to close the unified view gap
- Data quality failure in early months damages trust permanently
- Enterprise customers may prefer CDP vendors (Segment, Salesforce) for data layer

ARCHITECTURE

How the Platform Works — System Overview

Five layers that turn raw product events into trusted customer insight

1	2	3	4	5
Data Sources	**Canonical Schema**	**Storage Tiers**	**Access Control**	**Customer Interfaces**
Sinch Products	Communication Event	Hot / Warm / Cold	5-Layer Security	Any Access Pattern
SMS API	15 event types	Hot: 0–90d · <3s	OAuth 2.0 + JWT	REST API + SDKs
Email (Mailgun)	8 schema blocks	Warm: 91–365d · <30s	5 RBAC roles	Dashboards + Alerts
Conversation API	Immutable append-only	Cold: 365d+ · 1–5min	Tenant isolation	Bulk Export (Parquet)
Voice API	Schema validation	Pre-aggregation tables	Audit log	Warehouse Connectors
Verification API	Versioned contracts	Daily materialized views	Data residency (EU/US)	AI / NLQ (Phase 2+)
RCS / WhatsApp				
Wave 1–3 integrations	*The normalization layer*	*Performance by use case*	*Trust by design*	*Meet customers where they are*

The canonical schema (Layer 2) is the most critical technical decision — it is what makes Layers 3, 4, and 5 possible at all.

Architecture Layer 1:

The Canonical Data Foundation

Why a single event schema is the most important technical decision on this platform

Design decision: Event-centric immutable append-only log · 15 canonical event types · Cross-product unified schema · Billing-native fields

Identity

event_id, account_id, product_id, channel — the hard isolation boundary

Event Lifecycle

event_type, timestamp, status — 15 canonical event types

Parties

sender_id, recipient_id_hash — PII-safe HMAC hash for cohort analysis

Delivery Intel

error_category, latency_ms, attempts — 8 canonical error categories

Engagement

opened, clicked, replied — unified engagement signals across channels

Billing

billable, cost, unit_price — invoice and events share the same data source

Compliance

retention_class, data_residency_region, gdpr_erasure_status

AI Hints

ai_query_hints in catalog — semantic layer for NLQ accuracy

Outcome: SMS + Email + WhatsApp events use the same schema. "delivery_rate" means the same thing everywhere. Billing and events are the same data, not separate systems.

API-First Design Principles

Five key decisions — and why each one matters for customers

ARCHITECTURE

REST (not GraphQL)

Why: Predictable, cacheable, universal. Every developer knows REST. Arjun's team can integrate without learning a new query language.

Trade-off: GraphQL's flexibility traded for ecosystem compatibility

URL Versioning /v1/

Why: Explicit, bookmarkable, proxy-friendly. A URL with /v1/ survives corporate network filters and is immediately debuggable.

Trade-off: Header versioning is cleaner but less visible

OAuth 2.0 Client Credentials

Why: M2M industry standard. Arjun's pipeline gets a service account. 60-min token TTL + revocation list covers both security and practicality.

Trade-off: API keys are simpler but lack scope enforcement

Cursor Pagination + “since=”

Why: The 'since' parameter enables true incremental loads. Priya's ETL runs nightly and pulls only new events — no full re-scan.

Trade-off: Offset pagination is familiar but breaks on live data

Bulk Export as First-Class API

Why: Async job → signed Parquet URL → SHA-256 checksum. 24-hr download window. The export is not an afterthought — it's the primary path for large data consumers.

Trade-off: Synchronous export is simpler but unusable at scale

Access Control: Trust by Design

Five-layer security model — enforced at the query executor, not application logic

SECURITY & COMPLIANCE

Layer 1	**Identity & Authentication** OAuth 2.0 Client Credentials · JWT Bearer tokens (60-min TTL) · Revocation list for immediate invalidation
Layer 2	**Account & Sub-Account Isolation** account_id injected into every query by the executor — not application logic. Architecturally impossible to cross-tenant.
Layer 3	**Role-Based Access Control (RBAC)** 5 built-in roles mapped to the 5 personas: Admin · Analytics User · API Consumer · Billing Viewer · Compliance Officer
Layer 4	**Resource-Level Permissions** Field-level billing masking (billing fields return null without billing:read scope). Data residency enforcement (EU/US at launch).
Layer 5	**Immutable Audit Log** Every data access — including AI queries — logged synchronously. 365-day retention. Customer-visible. AI queries log the natural language input.

5 ROLES

Account Admin
IT / Security

Analytics User
Maya · Leila

API Consumer
Arjun · Priya

Billing Viewer
Stefan

Compliance Officer
Stefan / DPO

Guarantee: Your data is your data. No other customer, no Sinch engineer without your knowledge, and no AI interface can access it.

Architecture Layer 4:

AI Integrity Framework

Three non-negotiable properties — all three must be satisfied before any AI feature ships

GROUNDED

Every answer is derived exclusively from data that exists in the platform's storage layer, within the authorized scope. The AI never invents, extrapolates, or interpolates data it cannot trace to specific stored records.

Implementation:

RAG architecture: LLM translates intent → structured QuerySpec → query executor → results → LLM formats response. Parametric model knowledge never supplies metric values.

EXPLAINABLE

Every AI answer comes with a transparent 'How I got this answer' section showing: data source, time range, filters, records analyzed, calculation, and any caveats or data gaps.

Implementation:

Provenance block in every response. Mandatory. Cannot be disabled. Stefan can verify billing numbers. Leila can cite the source in board meetings.

ACCESS-CONSISTENT

The AI interface enforces the same access boundaries as the direct API — no more, no less. The AI is a query generation interface. It has no special permissions.

Implementation:

LLM outputs a QuerySpec (structured object, not raw SQL). QuerySpec is evaluated by the same query executor as direct API calls — including mandatory account_id injection.

Prioritization: What We Build and Why

Modified RICE framework + prerequisite gates + persona coverage balance

PRIORITIZATION

RICE Score =

(Reach × Impact × Confidence) ÷ Effort

Filter 1: Prerequisite Gate

Technically impossible features auto-defer regardless of RICE

Filter 2: Persona Coverage

Phase 1 must serve Maya + Arjun deeply, Stefan adequately

Phase 1 Committed Features

- Canonical event schema + 3-product integration
- Delivery / Volume / Error dashboards
- OAuth 2.0 + RBAC + multi-tenant isolation
- Billing reconciliation report + CSV/Parquet export
- Events API + Metrics API + Bulk Export API
- GDPR erasure + Audit log + Data residency (EU/US)
- Python + Node.js SDKs
- Data catalog with AI query hints

Explicitly Deferred (with reasons)

Alert engine

Needs 60-day stable baseline first — false alerts cause churn

NLQ AI interface

Phase 1 builds prerequisites; AI on bad data = worse than no AI

Fivetran connector

Needs stable v1 API for certification (Phase 2, Month 11)

Campaign dashboard

Needs campaign taxonomy standardized across all products first

Roadmap: Outcome-Based View

Direction and intent — what we commit to, in what order, and why

ROADMAP

NOW

Months 1–6 · Build a foundation customers can trust

Unified Data Layer

3 products integrated · ±0.5% accuracy guarantee · Immutable schema

Secure Access

OAuth 2.0 · 5 RBAC roles · Hard tenant isolation · GDPR-ready

API-First Access

Events + Metrics + Bulk Export APIs · Python + Node.js SDKs · 30-min quickstart

Operational Visibility

Maya's delivery dashboard · Scheduled reports · Error breakdown

Financial Reconciliation

Stefan's billing report · CSV + Parquet export · Audit log

NEXT

Months 7–12 · Become the operating system for Sinch data

Complete Channel View

Voice + Verification APIs integrated · 6+ products contributing data

Cross-Channel Intelligence

Multi-channel overview · Campaign performance · Executive dashboard

Proactive Alerts

Delivery rate drop alerts · Volume anomalies · Slack + Webhook delivery

Warehouse Integration

Fivetran connector GA (Month 11) · Airbyte secondary option

AI Interface Beta

NLQ for 20–30 design partner accounts · Full explainability block

LATER

Months 13–24 · From data reporting to intelligence

Anomaly Detection

Proactive alerts with causal analysis · 'Why did this happen'

AI GA

NLQ for all Analytics Users · Prescription recommendations

Benchmarking

Privacy-preserving cross-account benchmarks · Industry percentiles

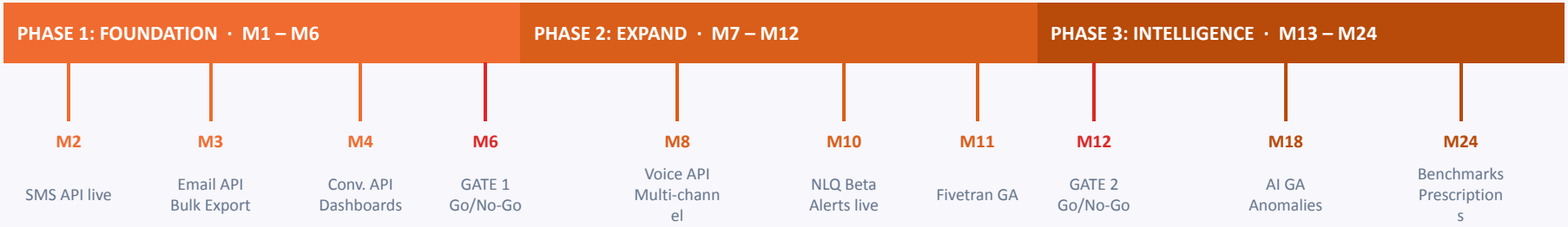
Ecosystem Expansion

Segment CDP · Additional warehouse connectors · Partner API

Roadmap: Timeline & Phase Gates

Milestone sequencing — with explicit go/no-go criteria before advancing

EXECUTION PLANNING



PHASE 1 GATE — ALL must be true

- Delivery rate accuracy ±0.5% · 30 consecutive days of validation
- ≥ 3 products live · completeness ≥ 99% · latency ≤ 5 min
- API v1: ≤ 0.1% error rate · p95 latency ≤ 500ms
- Zero cross-account data access incidents
- ≥ 5 design partner accounts active (first WDAA cohort)
- Developer onboarding: first API call ≤ 30 minutes (median)

PHASE 2 GATE — ALL must be true

- WDAA ≥ 40% above Phase 1 gate baseline
- ≥ 6 product teams contributing validated data
- Alert false positive rate ≤ 5%
- Fivetran connector certified · ≥ 3 accounts syncing
- NLQ accuracy ≥ 88% on test query set (100 queries)
- API v1: 99.9% uptime over trailing 90 days

OKRs: H1 — Build a Foundation Customers Can Trust

Q1 + Q2 (Months 1–6) · Calibrated for 70–80% attainment with strong execution

OKRS · H1

O1: Ship a trustworthy, unified data layer

- KR 1.1: Delivery rate accuracy ±0.5% vs. source truth · 30 consecutive validation days
- KR 1.2: Event completeness ≥ 99% across all Wave 1 integrations (rolling 7-day)
- KR 1.3: All 3 Wave 1 integrations complete by Month 4 · Wave 2 begins Month 5

O2: Developer experience that earns trust before the first dashboard

- KR 2.1: Median time-to-first-API-call ≤ 30 min · measured in ≥ 10 design partner sessions
- KR 2.2: Developer docs accuracy ≥ 4.2/5.0 · ≤ 3 'I got stuck' moments per session
- KR 2.3: Bulk export completion ≥ 99.5% · p95 completion ≤ 15 min (90-day history)

O3: Deliver the operational reporting experience that replaces the Friday spreadsheet

- KR 3.1: ≥ 60% of design partner accounts complete ≥ 2 meaningful actions in first 14 days
- KR 3.2: ≥ 40% of activated accounts have ≥ 1 active scheduled report by Month 6
- KR 3.3: ≥ 70% of Stefan-type accounts download ≥ 1 billing report within 30 days

O4: Build a secure, compliant platform enterprise legal teams approve

- KR 4.1: Zero cross-account data access incidents in Phase 1 production
- KR 4.2: ≥ 90% of design partner enterprise accounts pass security review without architectural changes
- KR 4.3: GDPR erasure verified by ≥ 1 DPO walkthrough · Audit log gap-free for 30 days

OKRs: H2 — Grow Adoption, Deepen Value, Begin Intelligence

Q3 + Q4 (Months 7–12) · North Star target: 500 WDAA accounts by Month 12

OKRS · H2

O5: Grow WDAA to establish platform momentum

- KR 5.1: WDAA ≥ 500 accounts by Month 12 (from ~25–40 at Phase 1 gate)
- KR 5.2: D30 retention ≥ 55% of accounts activating in H2
- KR 5.3: ≥ 30% of accounts have ≥ 2 distinct active users by Month 12

O6: Complete the channel view and unlock multi-channel analysis

- KR 6.1: ≥ 6 product teams with active validated integrations by Month 12
- KR 6.2: ≥ 40% of WDAA accounts query data from ≥ 2 channels in trailing 30 days
- KR 6.3: ≥ 35% of WDAA accounts have ≥ 1 active configured alert

O7: Establish the warehouse integration pathway for data engineers

- KR 7.1: Fivetran connector certified · ≥ 10 accounts actively syncing to warehouse by M12
- KR 7.2: Median warehouse connector time-to-first-sync ≤ 2 hours
- KR 7.3: Bulk Export p99 ≤ 20 min at 10× H1 volume (scale validation)

O8: Launch AI interface beta and establish a learning foundation

- KR 8.1: NLQ beta with ≥ 20 accounts by Month 10 · ≥ 60% make ≥ 5 queries in first month
- KR 8.2: NLQ interpretation accuracy ≥ 88% on 100-query test set at Month 12
- KR 8.3: Zero access control violations attributable to AI interface in H2

What Success Looks Like

Concrete moments — not metrics — that tell us the platform is working

Maya

Month 4

"Opens Monday dashboard and sees SMS, email, and WhatsApp delivery rates from last week — in one view, with consistent definitions. She doesn't open a spreadsheet. The number she presents on Friday is the same number she checked on Monday."

Arjun

Month 3

"Calls the Events API with one key, paginates 90 days of SMS delivery data with a cursor, downloads a Parquet file with SHA-256 verified. First query: 22 minutes after account provisioning."

Stefan

Month 4

"Downloads a billing reconciliation report that matches the invoice line by line. No support ticket. No 3-day wait. Cross-references it in 8 minutes."

Priya

Month 11

"Runs one Fivetran command. Sinch data starts syncing into Snowflake daily. Schema in the warehouse matches the catalog exactly. She didn't write any ETL code."

Leila

Month 9

"Gets a platform-generated quarterly summary in her inbox every Monday. The numbers match what the team has been reporting all quarter — no reconciliation meeting needed."

The Critical Internal Dependency

The roadmap that external customers see depends on internal product team integration

EXECUTION RISK

SWOT finding: Internal product team resistance is the highest-probability threat to the platform's success — not technology, not competition.

Wave 1 Months 1–4	**SMS API · Email API · Conversation API** Highest volume · covers 3 core channels · proves canonical schema handles structural diversity (threads vs. single messages)	*Risk: Medium — requires PM-level integration sponsors per team + signed integration contracts*
Wave 2 Months 5–8	**Verification API · Voice API** Verification: simple event schema, fast integration. Voice: CDR mapping is the most complex canonical mapping in the roadmap.	*Risk: Medium-High for Voice — CDR schema requires careful canonical mapping work*
Wave 3 Months 7–12	**RCS API · WhatsApp Direct · Additional products** Once platform value is demonstrated, remaining teams integrate via pull rather than push. Integration becomes self-serving.	*Risk: Low — by this point the platform has proven value and teams want in*

Critical path: Schema → Wave 1 (M4) → 60-day baseline → Alert engine (M9) → NLQ (M10). Any Wave 1 slip cascades to every downstream deliverable.

Let’s Build It!

This presentation summarises 14 separate documents prepared across strategy, architecture, roadmap, and execution workstreams. A working interactive prototype has also been built and shared separately. Happy to walk through the backbone documents on request.