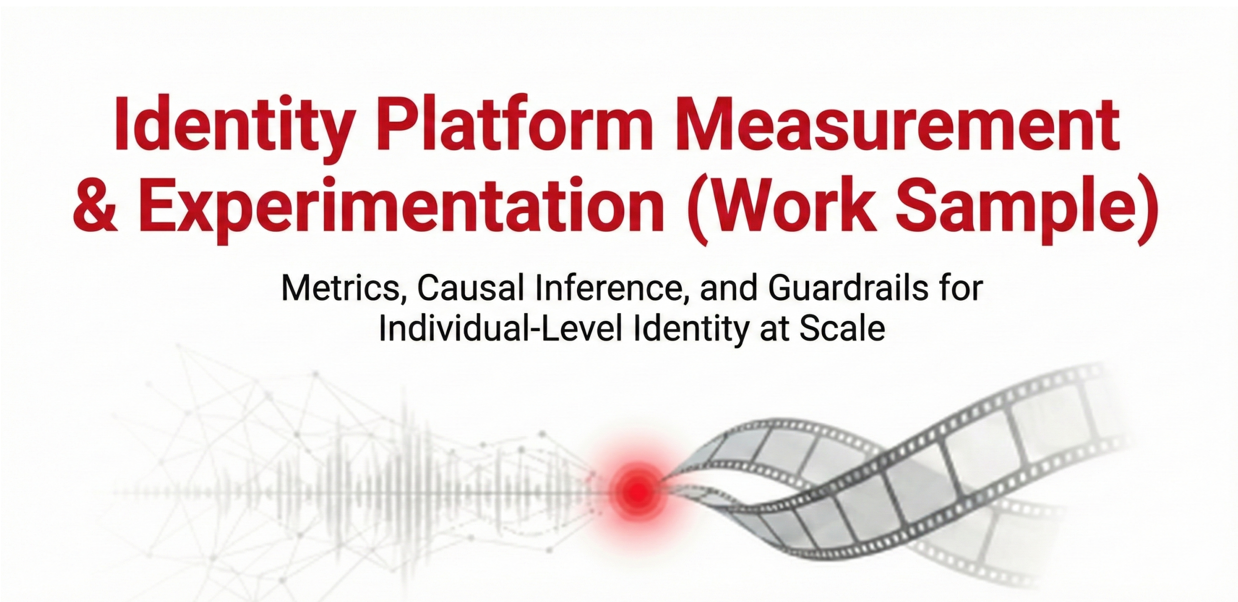
Identity Platform Decision Science — Work Sample (Hiring Memo)

Target role: Data Scientist L4/L5 — Identity DSE (Commerce)

IRIS YAHSUAN YANG

Los Angeles, CA • (626) 551-8382 • irisyang886@gmail.com • linkedin.com/in/irisyyang



Role fit: Demonstrates metrics research, experimentation and observational causal inference, and governance-minded operating models for building a unified, individual-level identity platform and measuring its downstream commerce and experience impact.

Notes on data & confidentiality: This memo repurposes methods from a larger public-data work sample (RetailRocket event logs) and uses synthetic illustrations for identity concepts. No proprietary data or internal systems are referenced. This document intentionally avoids brand logos/wordmarks.

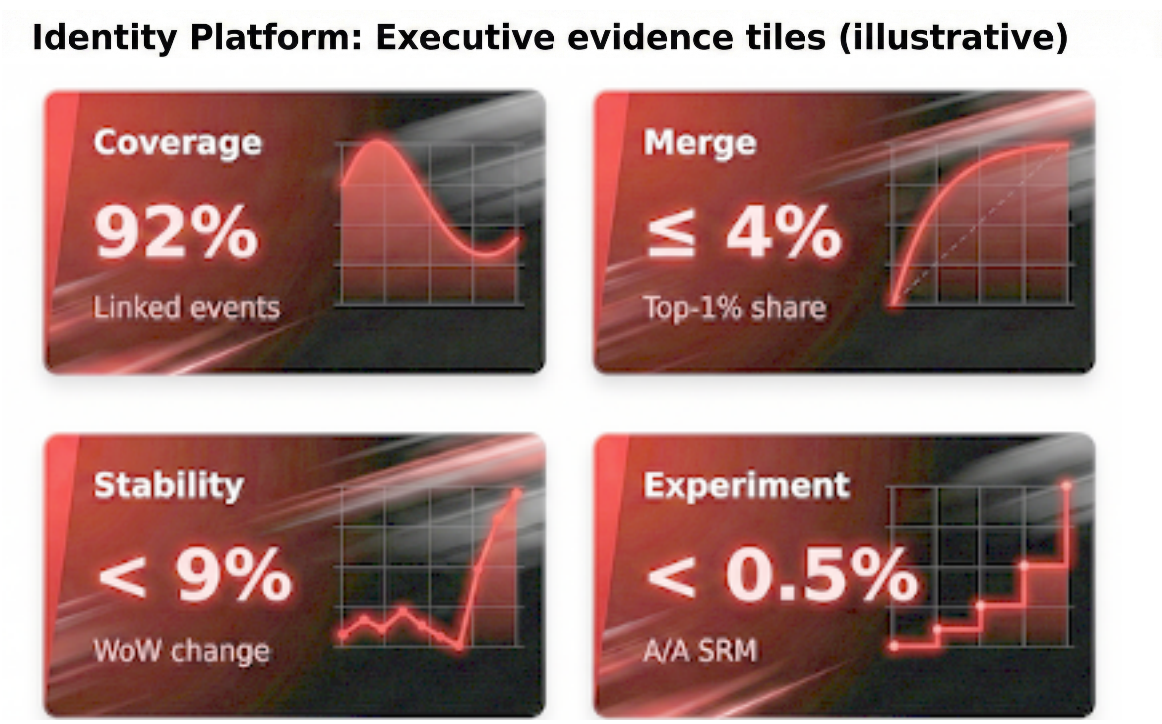
Abstract: Identity is a foundation for commerce decision-making—attribution, experimentation, personalization, and lifecycle measurement all degrade when identity is fragmented or incorrectly merged. This memo proposes a decision-grade measurement and experimentation framework for an identity platform: how to define success (coverage, correctness, stability, governance), how to validate impact with experiments and disciplined causal inference, and how to operate the system safely through observability, drift monitoring, and rollback criteria.

Executive summary

The identity platform's job is not only to connect identifiers—it is to create a stable unit of analysis that downstream teams can trust. In commerce and experience contexts, identity quality directly affects attribution denominators, experiment assignment, and the ability to measure long-run value. The framework below organizes identity work the way decision-makers consume it: define what 'healthy' means, ship changes behind holdouts and ramps, and continuously monitor merge/split risk and stability.

Key takeaways (decision-first):

- Treat identity as an operating policy, not a one-time model: every change must be versioned, measurable, and reversible.
- Use a metric hierarchy: primary outcomes (attribution and downstream lift), fast leading indicators (coverage and continuity), and strict guardrails (over-merge, over-split, stability drift, privacy).
- Design experiments for feasibility: some 'true' outcomes are low base-rate, so leading indicators and variance reduction are required to learn on practical timelines.
- When randomization is infeasible, use observational causal inference as a complement—with overlap, balance, and sensitivity diagnostics—rather than relying on naive before/after readouts.



Illustrative placeholders; the memo focuses on decision frameworks transferable to first-party identity telemetry.

Figure 1. Executive evidence tiles for an identity platform (illustrative).

1. What decisions a unified identity platform enables

A unified, individual-level identity platform becomes the shared key that connects product, marketing, and finance. When identity is fragmented, the organization double-counts people (inflated reach, unstable denominators), loses cross-device continuity, and misattributes the effect of campaigns or product changes. When identity is incorrectly merged, the platform creates collisions that can distort personalization and measurement and can introduce privacy and governance risk. The Data Scientist's role is to make these tradeoffs explicit, measurable, and operationally safe.

In practice, identity decision science supports four recurring decision classes:

- Measurement integrity: define 'person' and 'household' denominators that stay stable over time for reporting and forecasting.
- Experimentation: assign and analyze experiments at the right unit (individual, account, household) and verify assignment integrity.
- Attribution and incrementality: separate real lift from selection effects when exposures are targeted or uneven across devices.
- Governance and platform health: monitor merge/split/stability drift, document changes, and enable fast rollback if guardrails break.

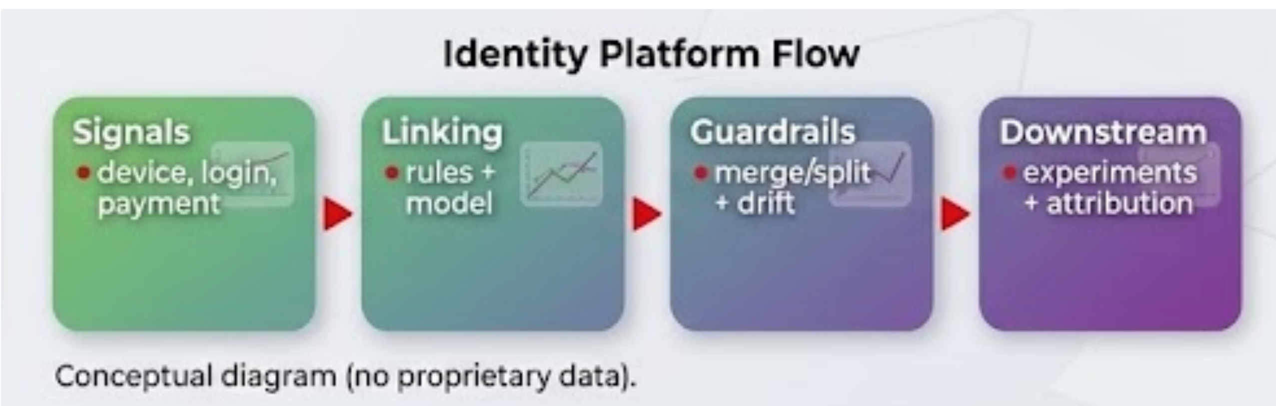


Figure 2. Identity platform pipeline: signals → linking → guardrails → downstream use (conceptual).

2. Metrics research: define success and guardrails

Identity platforms succeed when they improve decision quality downstream and remain safe to operate. That requires a metric hierarchy that is consistent across dashboards, models, and experiments. The hierarchy below mirrors the approach used in the broader work sample: outcomes for value, leading indicators for speed, and guardrails for safety.

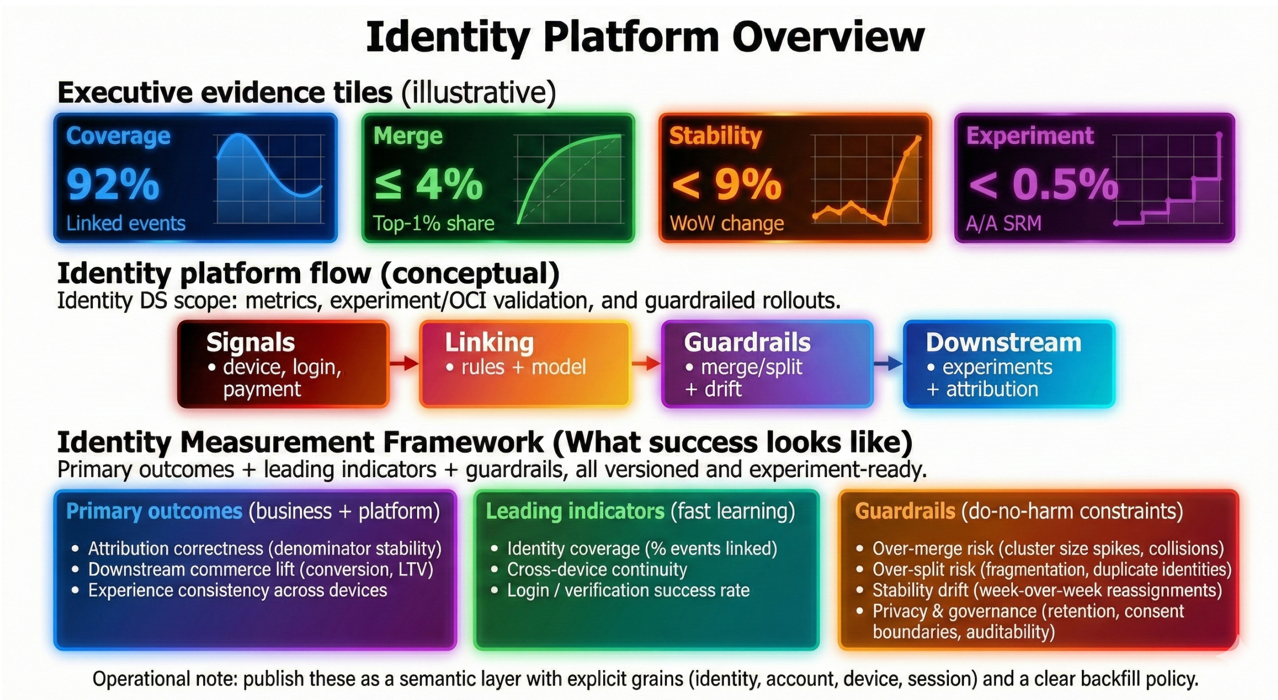


Figure 3. Identity measurement hierarchy: outcomes, leading indicators, and guardrails.

Coverage can be defined as the share of events that can be linked to a stable identity within an acceptable latency window. Correctness is the probability that linked identifiers truly belong to the same individual (precision), while fragmentation reflects the degree to which one person appears as multiple identities (recall). Stability measures how often identities change membership week-over-week absent true user behavior shifts. These definitions must be published with explicit grains (identity, account, device, session) to prevent denominator drift.

3. Guardrails: manage merge/split risk and stability drift

Linking decisions are inherently a tradeoff between coverage and correctness. Pushing the system to link aggressively can inflate coverage but increases over-merge collisions; being too conservative protects precision but can leave identities fragmented, degrading cross-device continuity and attribution. A decision-grade platform makes this tradeoff observable and chooses an operating point based on downstream costs.

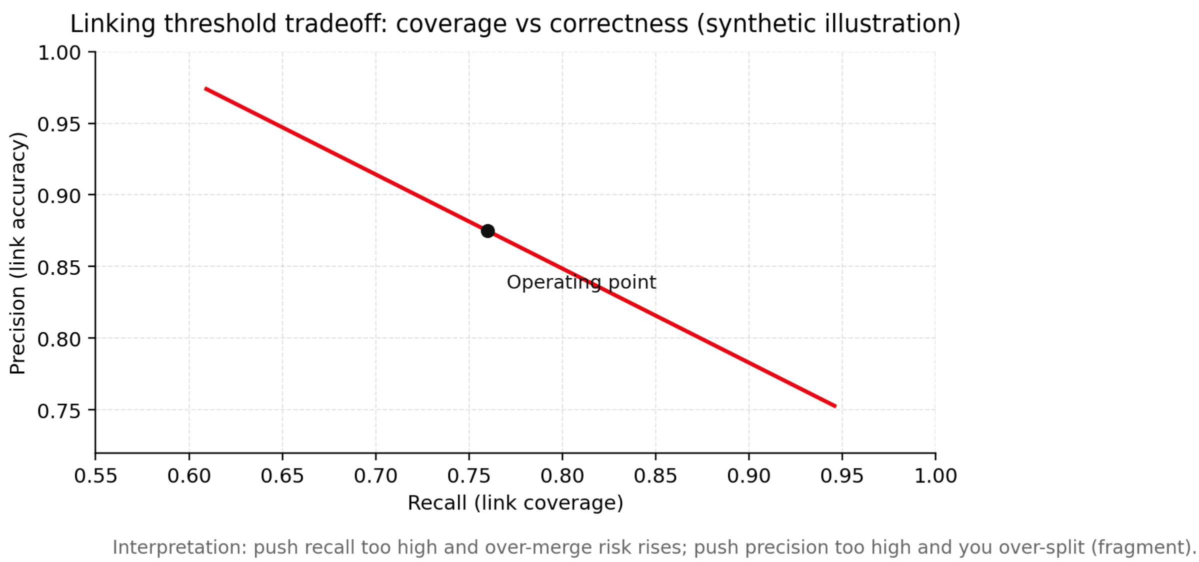


Figure 4. Threshold tradeoff (synthetic): recall/coverage vs precision/correctness; guardrails pick an operating point.

Stability monitoring is the early-warning system that catches regime changes: instrumentation shifts, model updates, or ecosystem changes that move the mapping even if headline coverage stays flat. In practice, stability is monitored overall and by segment (geo, device type, market), with alert thresholds and a clear on-call response: investigate, halt ramp, or roll back.

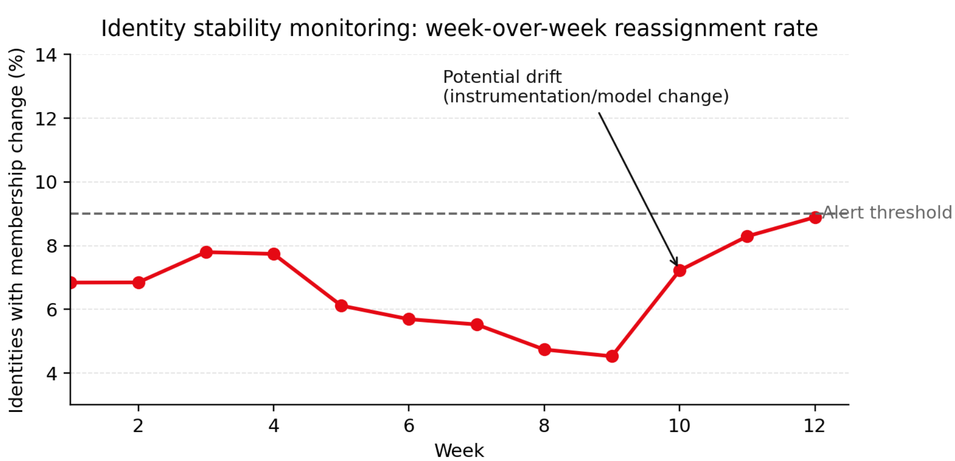


Figure 5. Stability Guardrail: Week-over-Week Identity Reassignment (Drift Monitor)

4. Experimentation-first impact validation (and feasibility)

Offline checks can catch obvious regressions, but only experiments establish incrementality. Identity changes should ship behind holdouts and staged ramps, with pre-registered primary metrics and guardrails. Feasibility matters: many downstream outcomes (purchase, long-horizon retention) can be low base-rate, which makes small lifts slow to detect unless the program uses leading indicators and variance reduction.

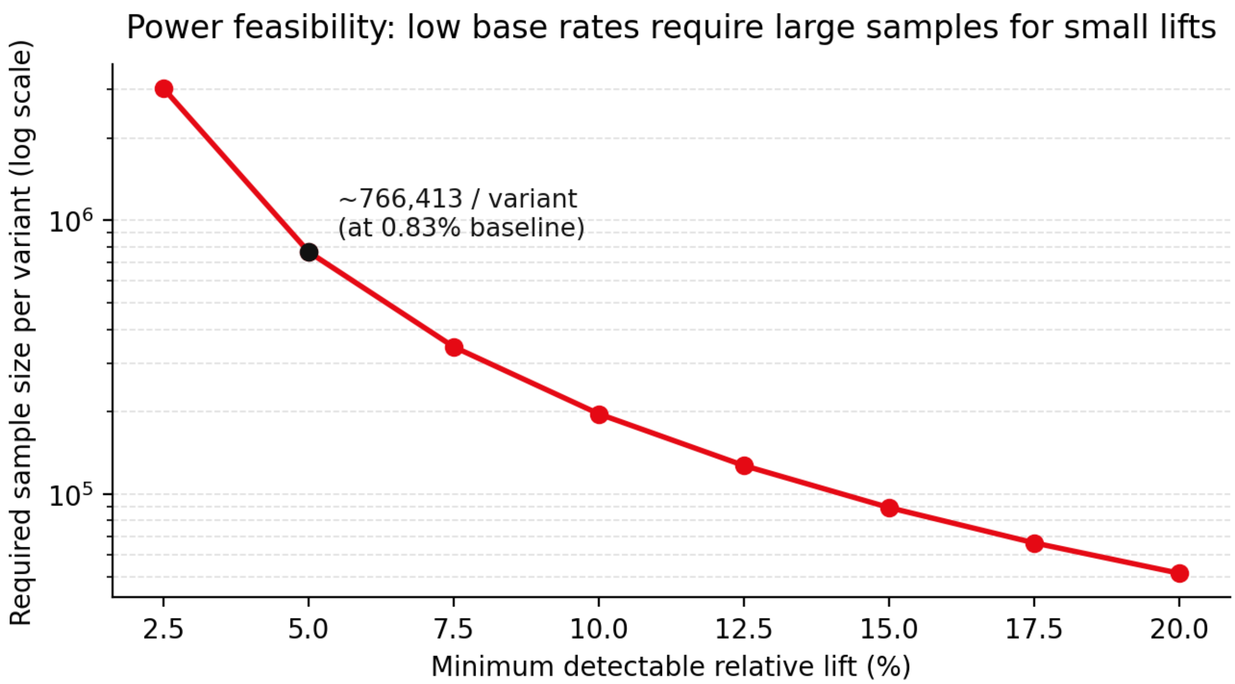


Figure 6. Power curve (proxy baseline): low base rates require large samples for small lifts.

For an identity platform, leading indicators often include changes in attribution denominator stability, cross-device session continuity, and reduced duplication in lifecycle targeting. Those can be measured at higher frequency than long-horizon value outcomes and used as primary metrics—while still keeping revenue/LTV as confirmatory endpoints.

5. When randomization is hard: observational causal inference as a complement

Identity rollouts are sometimes constrained: legal/privacy requirements, operational urgency, or system-level migrations can make randomization partial or impossible. In those cases, the organization still needs defensible estimates of impact. The safe approach is to treat observational causal inference as a complement to experimentation, not a replacement: use it to prioritize hypotheses and evaluate constrained rollouts, with explicit assumptions and mandatory diagnostics (overlap, balance, and sensitivity).

Selection Bias Inflates naïve lift; weighting moves estimate toward truth.

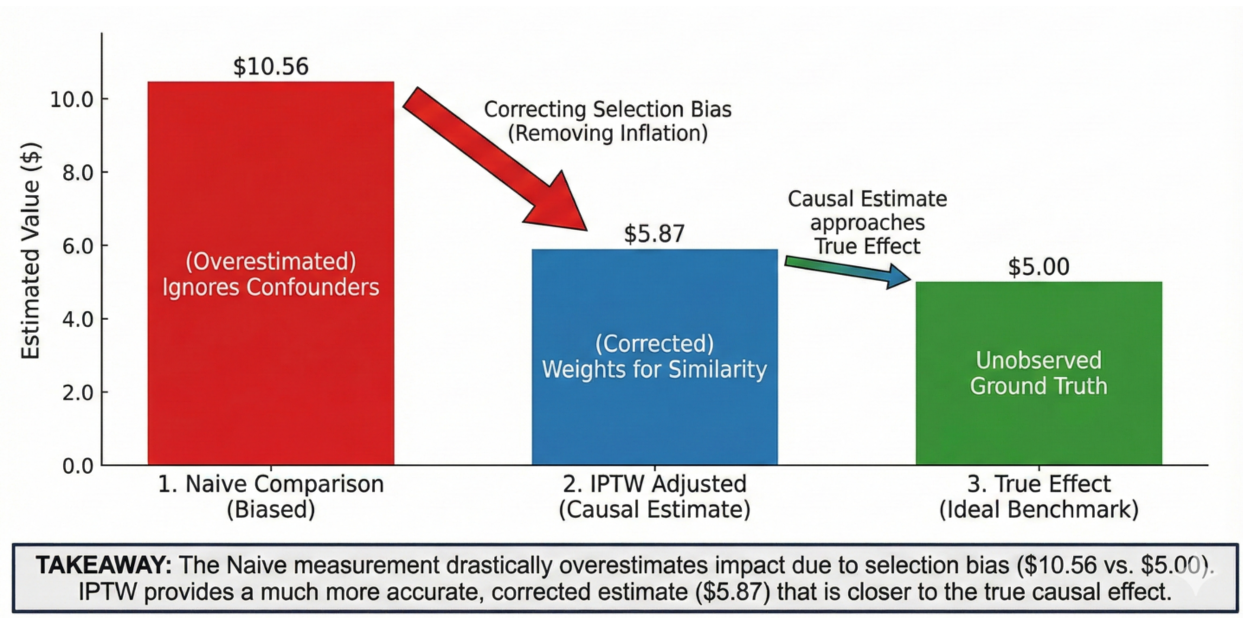


Figure 7. Selection bias demo (synthetic): naive comparisons can overstate lift; IPTW moves the estimate toward truth under assumptions.

Operationally, OCI is only as credible as the instrumentation. At minimum, you need (1) reliable exposure or assignment logs, (2) a clear temporal ordering so features precede outcomes, and (3) recurring balance checks so stakeholders can see whether the adjustment actually reduced confounding.

6. Data foundations: event taxonomy, semantic layer, and governance

Identity work collapses quickly without consistent definitions. The most common failure mode is not modeling—it is definitional drift: ‘identity,’ ‘account,’ and ‘household’ mean different things in different dashboards and experiments. A production-ready approach publishes a governed semantic layer with explicit grains and a clear backfill policy.

SIGNAL FAMILY EXAMPLES & IMPORTANCE

Signal family	Examples (and why they matter)
Identifiers	device_id, profile_id, account_id, household_id; enables deterministic join and experiment units
Authentication	login, success, verification, anchors high-precision links and reduces fragmentation
Payments / commerce	payment_token, purchase events; improves attribution and LTV measurement (governed)
Engagement	sessions, plays, clicks; supports continuity metrics and downstream modeling
Exposures	impressions by surface + rank; required for causal validity and OCI
Governance	consent flags, retention windows, schema versioning; enforces privacy-by-design and auditability

Data contracts and observability make identity safe to operate: freshness SLAs, join-rate monitoring across producers, schema/version change tracking, and anomaly alerts for sudden shifts in coverage or cluster-size distributions.

7. Calibration plan: connect identity changes to real outcomes

A platform only earns trust when it can explain what a change means for downstream decisions. Calibration is the process of mapping identity quality metrics to business outcomes under time-respecting windows, so that stakeholders can interpret changes consistently and avoid 'metric whiplash' after migrations.

Operational goal: Changes count as wins only if guardrails stay healthy while outcomes improve.

Figure 8. Calibration methodology: outcomes, windows, validation, and governance for versioned release.

A key principle is robustness: decisions should not flip based on small calibration uncertainty. That implies routine sensitivity analysis (scenario bounds, negative controls, placebo periods) and drift monitoring by segment so recalibration is a governed process, not an ad hoc fix.

8. Operating model: health dashboards, experiment observability, and rollbacks

Identity platforms are long-lived and constantly changing. The operating model should make rollouts safe: one command view that links platform health, experiment readouts, and data reliability so teams can diagnose issues quickly and roll back when guardrails break.

Figure 9. Operating model wireframe: identity health, experiment observability, and data quality in one view.

9. 30/60/90-day plan for a founding identity data scientist

Below is a practical execution plan that prioritizes decision leverage and measurement integrity. It is designed to be useful even while the platform is still being built.

Days 0–30: establish measurement and guardrails.

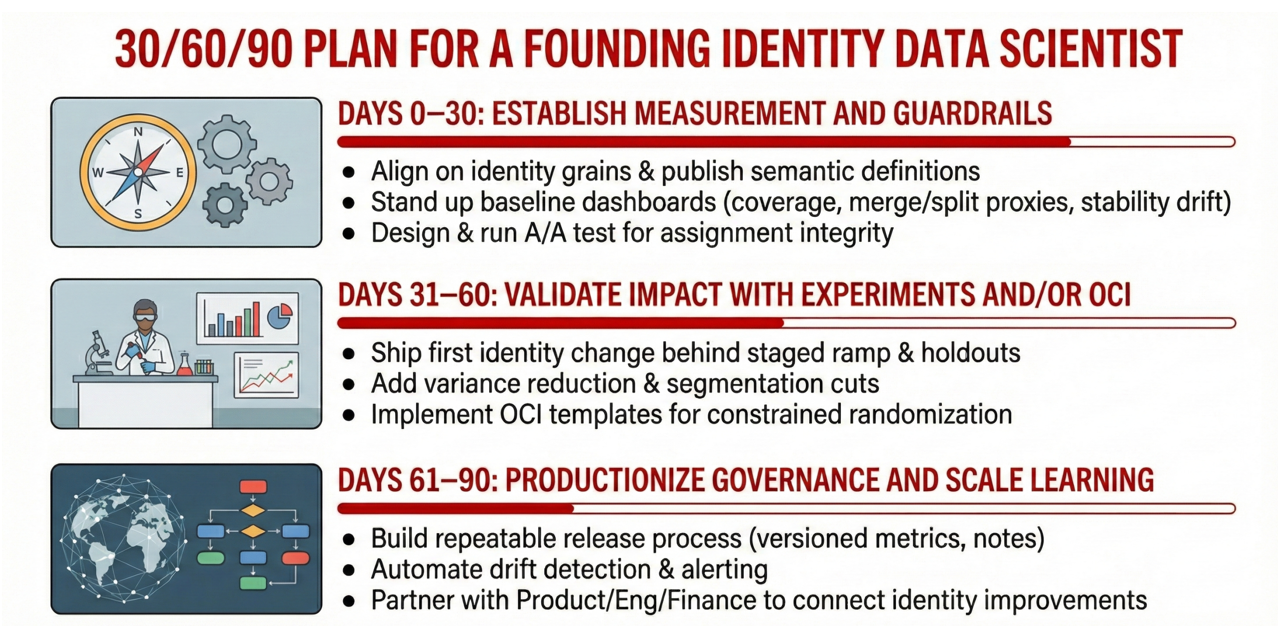
- Align on identity grains and publish the first semantic definitions (individual, account, household) with owners and backfill rules.
- Stand up baseline dashboards for coverage, merge/split proxies, stability drift, and data freshness/join-rate observability.
- Design and run an A/A test (or shadow evaluation) to validate assignment integrity, SRM checks, and metric stability before shipping changes.

Days 31–60: validate impact with experiments and/or OCI.

- Ship the first identity change behind a staged ramp with holdouts; pre-register primary metrics, confirmatory outcomes, and rollback triggers.
- Add variance reduction and segmentation cuts to improve feasibility for low base-rate outcomes.
- When randomization is constrained, implement OCI templates with required diagnostics (overlap, balance, sensitivity).

Days 61–90: productionize governance and scale learning.

- Build a repeatable release process: versioned metrics, release notes, experiment registry linkage, and a standard postmortem template.
- Automate drift detection and alerting; define retraining/recalibration triggers and ownership.
- Partner with Product/Engineering/Finance to connect identity improvements to downstream commerce strategy and roadmap decisions.



Appendix: What this memo reuses from the larger work sample

This hiring memo is adapted from a longer, public-data work sample focused on building an end-to-end measurement and decision system. The most directly transferable components for Identity DSE are: (1) a governed metric hierarchy with guardrails, (2) time-respecting evaluation and experiment feasibility (power and leading indicators), (3) observational causal inference discipline, and (4) system blueprints for semantic layers, observability, and experiment registries.

Key artifacts reused conceptually include the experimentation/OCI sections and the production blueprint pattern (metric layer → monitoring → causal learning loop).