Case Study: Address & Pincode Extraction from Company Websites (Senior Program Manager)

Why are you doing this case study?

A Senior Program Manager in the Product team is responsible for managing a diverse range of programs supporting the Tracxn Product Operations. These include, but are not limited to, Platform Abuse monitoring & detection, AI enabled Data Production (we are a data company), 3rd party Vendor management, and more. In this case study, we will use an example of AI enabled Data Production.

Problem Context

Tracxn needs to build a production system that extracts and maintains accurate address information (including pin codes) from company websites. This system must process both batch updates and real-time requests, ensuring our database always reflects the most current company location data.

Unlike a purely technical case, this requires program leadership across engineering, data science, and business operations. The Senior Program Manager must design the program strategy, governance framework, and cross-team execution model to ensure successful delivery.

Business Requirements

Triggers

1. Batch Processing: >100K companies (monthly refresh)
2. Usage-based Processing: >1K companies per day (on-demand updates)

Quality Standards

- High precision (errors have high business cost)
- Multiple office locations per company
- Data conflicts must be flagged for review

Role Expectations (Senior Program Manager)

As an SPM, you are not writing the regex or LLM prompts yourself. Instead, you ensure that:

- Cross-functional ownership (Data Engineering, AI/ML, Infra, QA, and Ops) is clearly defined.

- Roadmap, dependencies, and risks are tracked and managed.
- Quality and cost metrics are aligned with business goals.
- Scalability and sustainability are considered from day one.
- Governance and feedback loops are in place for continuous improvement.

Program Design

1. Workflow & Ownership Model
   - Discovery & Extraction
   - Data Engineering team: heuristics and crawlers for page discovery.
   - AI/ML team: LLM-powered extraction and model tuning.
   - Data Processing & Normalization
   - Shared responsibility: Data Ops for annotation, Engineering for normalization rules.
   - Conflict Resolution
   - Automated checks (engineering) + Manual review (Ops).
   - Database Integration
   - Infra team ensures deduplication, logging, and audit trails.

➡️ SPM responsibility: Clarify interfaces, and create a delivery plan (e.g., phased rollout: MVP → pilot → scale).

2. Measurement Framework
   - Precision & Recall
   - Defined jointly by Product + Data Ops, validated against ground truth samples.
   - Cost Efficiency
   - Per-million-record API cost tracked; dashboards monitored by Finance + Eng.
   - SLAs
   - Batch processing SLA (monthly completion within X hours).
   - On-demand SLA (<5s response for 95% queries).

➡️ SPM responsibility: Drive alignment on acceptable thresholds, ensure reporting is automated, escalate deviations early.

3. Governance & Communication
   - Periodic reviews with senior leadership (Product, Data, Finance, Infra).
   - Metrics Reporting: Precision/recall dashboards for leadership visibility.
   - Change Management: Ensure schema or pipeline changes are communicated across teams.

What's Expected from Case Study Submission

1. Program Plan
   - Clear phasing: MVP → Pilot → Scale
   - Roles, responsibilities, and handoffs

2. Metrics Framework
    - Precision, recall, cost, SLA – tied to stakeholders
3. Governance and communication
    - Metrics reporting, Change management, communication