Adaptive, Cost-Constrained LLM Governance Framework for Municipal Permit Data under Schema Drift and Regulatory Change

Jiangwei Wang
CIS 6930 (Spring 2026)
UFID: 51768605

February 21, 2026

1 Introduction and Motivation

Municipal open data systems, like the Building Permits dataset from the City of Gainesville, are frequently utilized for public transparency, regulatory monitoring, infrastructure forecasts, and economic planning. Despite their significance, open civic datasets have several issues that make longitudinal analysis and data governance difficult, including schema drift, changing category taxonomies, missing or distorted fields, and inconsistent lifecycle semantics.

With adaptive semantic normalization, anomaly explanation, and cross-temporal consistency, recent developments in large language models (LLMs) offer new ways to enhance conventional rule-based ETL workflows. Cost limitations, automation limits, schema evolution resilience, and governance-aware architectural design are still unanswered issues, nevertheless.

For the purpose of managing municipal permit data, this research suggests an LLM-enhanced, cost-constrained MCP ETL pipeline that is specifically assessed in terms of performance, robustness, cost effectiveness.

2 Dataset and Problem Scope

2.1 Data Source

- **City of Gainesville — Building Permits** (DataGNV; Socrata API).
- Accessible at: https://data.cityofgainesville.org/Building-Development/Building-Permits/p798-x3nx
- Dataset spans multiple years and contains permit categorization, dates, contractor information, status fields, and optional free-text fields.

2.2 Scope

The project focuses on:

- Semantic normalization of permit types across time and schema variants.

- Anomaly detection (lifecycle inconsistencies, missing fields).
- Cost-constrained inference strategies using LLMs.
- Architecture design for governance-aware municipal ETL.
- Evaluation across five research dimensions: Performance, Robustness, Efficiency.

Out of scope: predictive modeling of future permits and external socioeconomic forecasting.

3 Project Direction Declaration

Direction A: LLM Pipeline over a New Data Source.

The primary focus is the design and evaluation of an LLM-orchestrated MCP pipeline applied to the Gainesville Building Permits dataset, which has not been previously explored in this course.

4 Research Questions

This project is drift-centered: treat taxonomy drift and schema evolution as the primary stress condition for municipal permit data pipelines.

4.1 Primary Question: Drift Robustness

Can an LLM-governed permit normalization pipeline maintain longitudinal category stability under progressive taxonomy drift better than rule-based baselines?

4.2 Secondary Question: Performance

Does LLM normalization improve Macro-F1 and long-tail recall compared to deterministic rule-based mapping?

4.3 Cost Question

What is the trade-off between token cost, latency, and stability improvement under explicit token budgets?

5 Methodology and Pipeline Design

5.1 Canonical Taxonomy Design

An initial taxonomy of 20–30 normalized permit categories will be defined with clear rules and descriptions. This taxonomy will be refined iteratively through stratified gold labeling.

5.2 Pipeline Architecture

The proposed pipeline comprises:

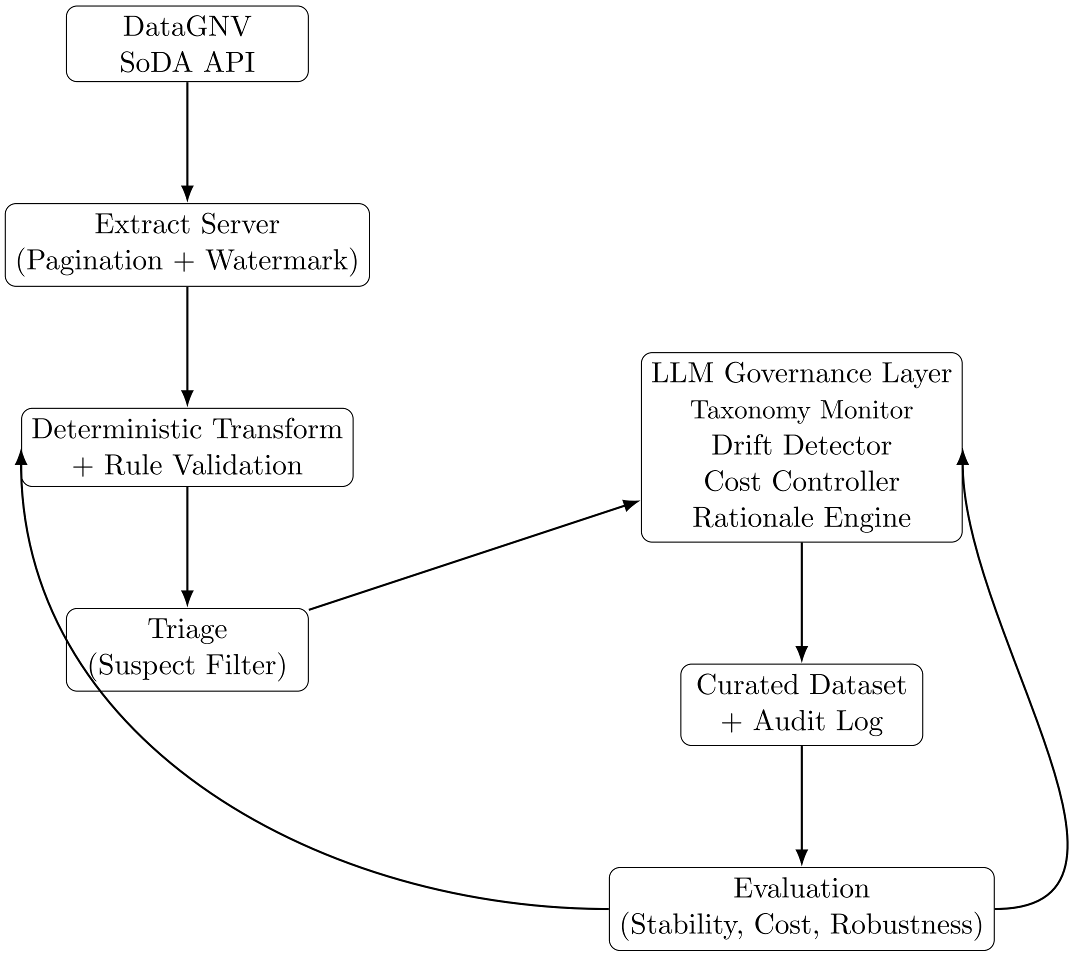

Figure 1: Adaptive Governance-Aware MCP Pipeline for Municipal Permit Data

Key components:

- **Extract Server:** incremental Socrata API pulls with pagination and watermarking.
- **Transform Server:** deterministic rules, baseline normalization, triage for suspect records.
- **Governance Layer (LLM Quality Inspector):**
  - taxonomy normalization with constrained prompts
  - anomaly detection and structured explanation
  - category evolution adaptation
  - cost controller and inference guardrails
- **Curated Output + Audit Trail:** final normalized table plus structured logs and rationale.
- **Evaluation & Reporting Server:** metrics, visualizations, cost summaries.

5.3 Cascade and Budget-Constrained Strategies

To limit cost while maintaining quality:

- Rule + statistic triage → LLM only on "suspect" records.
- Batching, caching, and threshold-based invocation controllers.
- Token budget optimization schedules for periodic runs.

6 Evaluation Plan

6.1 Gold Standard Labeling

- **Phase 1:** 500 stratified records to refine taxonomy and prompts.
- **Phase 2:** Expand to 1,000–2,000 records for final evaluation.

Each label contains: `canonical_permit_type`, `canonical_status`, anomaly tags, and (if applicable) the correct patch.

6.2 Evaluation Metrics (Permit Governance-Specific)

Task	**Concrete Metrics Used in This Project**
Permit Type Normalization	Macro-F1 across 20 canonical categories; Long-tail Recall (bottom 25% frequency classes)
Lifecycle Anomaly Detection	Precision, Recall, F1; False Positive Rate
Schema Drift Robustness (Primary)	Degradation curve under increasing drift intensity; Relative Accuracy Drop (%); Category Stability Index (CSI)
End-to-End Efficiency	Throughput (records/sec); Latency per 1k records; Token cost per 1k records; Cost per 1% CSI improvement

Table 1: Concrete Evaluation Metrics for the Drift-Centered Permit Governance Study

6.3 Controlled Drift Experiments

To rigorously evaluate robustness and cost trade-offs, we simulate progressive taxonomy drift under controlled regimes. Drift intensity varies from 0% to 40% (in 10% increments).

Experiment	**Measurement**
Category Renaming Simulation	Randomly rename 10%–40% of permit type labels; measure degradation curve
Category Split/Merge Simulation	Split one class into subclasses or merge related classes; measure stability recovery (CSI)
Noise Injection Test	Inject spelling noise and date perturbations; compute robustness degradation
Budget-Constrained Sweep	Token budget $B \in \{0, 1k, 5k, 10k\}$; plot Cost–CSI frontier
Long-Tail Stress Test	Evaluate recall on rare permit categories (bottom quartile frequency)

Table 2: Controlled Drift and Cost Experiments (Drift-Centered)

Category Stability Index (CSI). CSI is defined as the Jensen-Shannon similarity between yearly permit category distributions before and after normalization. CSI captures longitudinal statistical stability, which is critical for policy and planning analyses under evolving administrative taxonomies.

6.4 Baseline Comparisons

- **Rule-Only Baseline:** deterministic regex + mapping tables.
- **Statistical Anomaly Baseline:** constraint-based checks with no LLM.
- **Similarity Embedding Baseline:** nearest neighbor string/embedding category mapping.
- **LLM-Only Upper Bound:** every record passed to LLM.

7 Risk Factors and Mitigation

- **Hallucination Risk:** Use constrained taxonomy prompts, rationale output, and confidence scoring. No blind overrides.
- **Token Budget Overruns:** Cascade strategies, batching, invocation policies.
- **API Rate Limits:** Retry policy, caching, incremental pulls.
- **Taxonomy Ambiguities:** Iterative refinement with stratified labeling feedback.

8 Project Milestones

- **Proposal (Feb 23):** Complete problem framing, architecture, evaluation plan.
- **Design Review (Mar 2):** Architecture detail, taxonomy draft, prompt sketched.
- **Code Checkpoint (Mar 23):** ETL + Transform + LLM Quality stub.
- **Draft Paper (Mar 30):** Preliminary results, error analysis.
- **Final Paper (Apr 13):** Full evaluation, figures, tables.
- **Presentation (Apr 20 week):** Demo and live Q&A.

9 Expected Deliverables

- GitHub repo with MCP servers, tests, and documentation.
- Curated normalized dataset + audit logs.
- Evaluation artifacts (scripts, summaries, visualizations).
- Draft and final paper (PDF).
- Presentation slides/demo video.

10 Conclusion

This project evaluates the stability of municipal permit data under progressive schema drift using an LLM-governed, cost-constrained MCP pipeline.

This research establishes the resilience threshold between adaptive LLM-based governance and deterministic rule-based systems by assessing stability indices, degradation curves, and cost-performance trade-offs while simulating taxonomy evolution.

The main contribution is an evaluation system that is drift-centered and prioritizes longitudinal data stability above isolated classification accuracy.