TECHNICAL CASE STUDY

Geospatial Systems AI Evaluator Framework

Turning four interface failure modes into testable enterprise design requirements

CLIENT	INDUSTRY	DURATION	TEAM
Independent portfolio	Enterprise GIS	Not documented	Keith Trowbridge

At a Glance

4	4.5:1	32–36	0
Review criteria	Normal text target	Row height px	Reloads target

Figure 1. Documented scope, reference values and design targets; labels are not verified production results.

1. Executive Summary

This independent portfolio study examines a conceptual real-time fleet dashboard in which visually polished AI-generated layouts conflict with operational use. Keith Trowbridge organizes the review around basemap legibility, marker aggregation, spatial query continuity, and enterprise data density. The proposed solution combines restrained cartography, scale-aware summaries, shared map-and-table state, and compact accessible records. The deliverable is a traceable evaluation and redesign framework rather than a verified production deployment or a measured revenue improvement. [SRC; W1–W3; G1–G2]

> KEY TAKEAWAY
> **Operational behavior requires evidence beyond visual polish.**

2. Background & Challenge

2.1 Context

The source scenario places a dispatcher in a logistics center monitoring thousands of changing asset variables. A useful interface must support exception discovery, geographic orientation, and comparison across records while positions and statuses change. ArcGIS Enterprise and Oracle Spatial appear as domain context; the documents do not establish that both platforms were installed or integrated. A researched reference design should fit an existing service boundary rather than require a new database simply to implement a better screen. [SRC; G1–G2]

Illustrative photograph. AI-generated conceptual setting; not an actual facility, client, product screenshot or delivered customer portrait.

The evaluator guide distinguishes a polished visual appearance from functional enterprise architecture. It emphasizes training signals and reward-model weighting, but supplies no annotation dataset or trained reward model. This version adds a review protocol: evaluators inspect the same scenario, identify criterion-level evidence, record severity and adjudicate disagreements before any label is used for model training.

Planning baseline. Use synthetic assets with stable IDs, positions, status, last-updated values and an explicit coordinate reference. Include sparse and dense distributions, repeated coordinates, isolated urgent assets and clusters containing one urgent member. The 5,000-asset and ten-session scenario in

the expanded source is useful as one test configuration; vary it rather than describing it as an actual customer estate. Record update cadence and viewport because both influence visibility.

2.2 The Problem

A high-detail raster competes with small labels and thin boundaries; fixed-size markers collapse into overlapping symbols at regional scales. Coordinate-entry forms and page reloads disrupt geographic reasoning. Oversized cards reduce simultaneous comparison, so a dispatcher has to remember off-screen values. These are credible failure modes illustrated by the source, not evidence that all generative models fail at a measured rate. [SRC]

- **Legibility: background texture and status symbols compete at similar visual weight, making exceptions harder to distinguish.**
- **Spatial scale: identical pin footprints obscure both regional distribution and individual urgent assets.**
- **Workflow continuity: nested filtering and a hard refresh separate map navigation from record review.**

2.3 Goals & Constraints

Goal	Success Criteria	Constraint
Readable layers	Normal text ≥4.5:1; relevant graphics ≥3:1	Evaluate actual backgrounds and states
Useful aggregation	Accurate counts and visible exceptions at each scale	Tune against viewport and symbol size
Persistent querying	Map center, scale and selection retained	Cancel or ignore stale responses
Comparable records	32–36 px desktop rows where usable	Allow wrapping, zoom and accessible controls

Table 1. Project objectives and constraints.

3. Solution Overview

The proposed architecture treats the map and table as two views of one authoritative selection. A low-detail basemap provides orientation; aggregate symbols communicate counts and urgency at wider scales. Drawing a boundary changes shared query state and updates related rows without resetting the camera. Density is used to improve comparison, with readable typography and controls taking precedence over a fixed row-height target. [SRC; G1–G2]

Basemap Flaw

Basemap Correction

Clustering Flaw

Clustering Correction

Source visual evidence. Basemap and clustering flaw/correction illustrations supplied in the archive; these are conceptual comparisons.

Cartographic hierarchy. Retain enough roads, boundaries and place labels to orient the dispatcher, but reduce decorative texture and saturation beneath active data. Treat selection, warning and critical states as named tokens with consistent meaning. A label backplate is often more dependable than darkening an entire imagery layer because it isolates the local background. Inspect hovered, selected, disabled and updating states, not just the default mockup. The 4.5:1 normal-text and 3:1 large-text requirements apply within their criterion scope; relevant non-text graphics have separate requirements. A full conformance review also needs keyboard, focus and other checks. [W1–W3]

3.1 Approach

1. **Discover: inventory dispatcher tasks, geometry types, data freshness and current interaction steps; capture comparable baseline states.**
2. **Design: specify label treatments, cluster counts, severity summaries, boundary semantics and non-drag alternatives.**
3. **Build: implement shared IDs, a versioned query adapter, compact records and explicit pending, empty and error states.**
4. **Validate and launch: run task trials and accessibility checks across map scales; pilot only after counts, queries and failure recovery agree.**

3.2 Technology Stack

Layer	Technology	Rationale
Presentation	Semantic HTML, CSS and JavaScript	Reference implementation; reuse approved shell
Map	ArcGIS Maps SDK or approved equivalent	Supported clustering and spatial queries [G1–G2]

Layer	Technology	Rationale
Data	Existing feature service; optional spatial adapter	One source of truth with geometry and timestamps
Evidence	Versioned scenarios and structured review records	Connect each judgment to a reproducible state

Table 2. Technology stack and selection rationale.

4. Architecture & Implementation

4.1 System Architecture

Actions change a shared state object containing camera, selected IDs, active geometry and the current request generation. An asynchronous spatial query returns matching features and record metadata. The reconciliation layer accepts only the latest valid result, then updates map highlights, counts and table rows together. Geometry and coordinate reference information remain authoritative in the data service; screen-space clusters summarize display density and do not become geographic query polygons. [G1–G2]

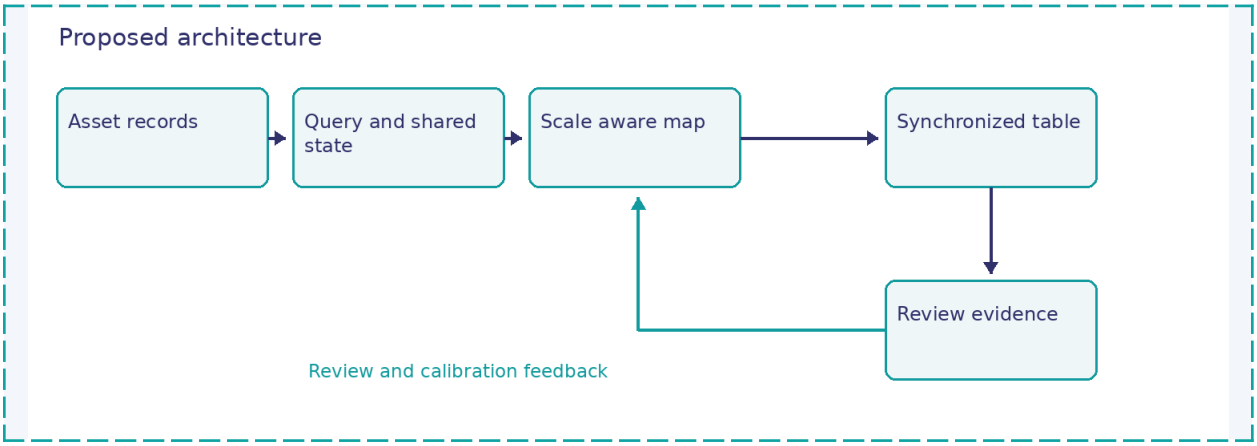

Figure 2. Proposed architecture and information flow; diagram is not deployment evidence.

Query semantics. Specify whether a boundary selects assets on its edge and whether the query evaluates latest or snapshot positions. Use a supported spatial reference and validate geometry. Cancellation or a request-generation guard prevents an old response from replacing a newer result. Service pagination and transfer limits must be handled before claiming map and table counts agree. A partial query must be visibly labeled rather than presented as a complete fleet result. [G2]

4.2 Key Technical Decisions

DECISION

Preserve context through asynchronous map-native filtering instead of submitting a coordinate form that reloads the page. The simpler form remains useful as an accessible equivalent or advanced input, but must drive the same state. Prefer screen-space clustering to a universal map-scale switch; the source's 1:50,000 value is an initial heuristic, not a cartographic standard. Trade-off: state reconciliation and accessible alternatives require more work than a static screenshot. [SRC; G1–G2; W3]

Priority preservation. An aggregate should expose total count and the number or severity of exceptions. A highest-priority color alone cannot reveal how many alerts are contained. Offer drill-down and a synchronized table; determine whether critical markers remain separate or receive a persistent overlay. Screen-space clusters change as the camera changes, so their geographic meaning must not be overstated. [G1]

4.3 Code or Configuration Snippet

```
// Illustrative integration logic, not deployed code
let generation = 0;
async function filterZone(geometry) {
  const current = ++generation;
  setPending(true);
  try {
    const q = assets.createQuery();
    q.geometry = geometry;
    q.spatialRelationship = "intersects";
    const result = await assets.queryFeatures(q);
    if (current === generation) commit(result);
  } catch (e) { if (current === generation) showError(); }
  finally { if (current === generation) setPending(false); }
}
```

Listing 1. Illustrative logic; integration, error recovery and tests remain implementation work.

4.4 Challenges Encountered

CHALLENGE AND RESOLUTION

The source suggests a 40–50% imagery overlay and compact rows as corrective measures. Neither is a universal pass condition. Contrast varies with the background; rows must expand for wrapped values and controls. Use opaque label backplates where necessary, test actual foreground/background pairs, and provide severity labels and icons in addition to color. Critical records must remain discoverable inside aggregates. [SRC; W1–W3; G1]

Freshness and reconciliation. Define how an asset update arriving during a spatial query affects the result. If selection is snapshot-based, show its observation time; if it is live, reconcile membership and

explain changing totals. Use immutable IDs rather than display names and preserve table selection when sorting. An empty result, a stale result and a failed request are different states with different recovery actions. A retry must not reset the camera or silently clear the operator's selection. Server authorization still controls which records are returned; a client-side filter is not an access-control boundary.

5. Results & Impact

Evidence status. The source establishes the problem, requirements or described service. No supplied experiment log establishes production uptime, revenue uplift, measured user satisfaction or completed browser acceptance. The table distinguishes source observations from proposed corrections; quantified examples are not substituted for missing client measurements.

Metric	Before	After or Target	Evidence or Change
Normal text	Unreadable labels described	≥4.5:1 target	No measured baseline [W1]
Regional display	Overlapping fixed pins	Counts and alert summaries	Proposed correction [G1]
Filtering	Hard reload described	0 reloads target	Verify camera and focus
Row height	60 px example	32–36 px proposal	40–46.7% less height
Business outcomes	Not supplied	Instrument pilot	No revenue or time claim

Table 3. Source observations and proposed outcomes with evidence status.

Proposed study. Compare equivalent task scenarios in counterbalanced order: locate an urgent asset, select assets inside a region, revise a boundary and compare statuses. Record completion time, missed exceptions, incorrect selection, filter-to-result latency and camera continuity. Report median task time and the distribution of errors; retain failures and abandoned tasks. A pilot can identify usability issues before a larger quantitative trial establishes reliable effect sizes.

5.1 Business Impact

- **Efficiency: direct geometry selection and a stable camera should reduce repeated navigation; measure task time rather than inferring it from screen appearance.**
- **Revenue and cost: reuse existing GIS services where practical; hosting, service licensing and implementation effort require a project-specific estimate.**
- **Experience: compare alert discovery, selection accuracy and perceived effort; a cleaner map alone does not establish operator satisfaction.**

> **VALIDATION STATUS**
>
> No attributable client testimonial or independently verified outcome was supplied. The case study is complete as a researched analysis; implementation and commercial claims require the specified project evidence.

6. Lessons Learned

What Worked	What We Would Do Differently
Review behavior alongside styling	Add realistic task trials before approving generated layouts
Use counts, text and priority summaries	Test hidden exceptions and cluster transitions
Increase comparison density carefully	Include zoom, wrapping, focus and touch checks

Table 4. Retrospective highlights.

Evaluator protocol. Capture scenario ID, scale, viewport, interaction path and the criterion under review. Record the exact observation, its operational implication and the required correction. Have a second evaluator independently rate ambiguous cases; retain disagreements for adjudication. A visually appealing output should not receive a passing label if counts, keyboard access or query continuity fail. This creates usable evaluation evidence without implying that RLHF training has already happened. [A1]

7. Next Steps

The roadmap below is a proposed progression, not a commitment or a report of completed releases. Prioritize evidence that resolves the greatest uncertainty before adding visual scope.

- Next 30–90 days: implement a synthetic pilot and publish task, count and accessibility evidence for each of the four criteria.
- Next six months: calibrate thresholds with operators; compare sparse, dense and incident-heavy scenarios.
- Longer term: turn reliable human judgments into a versioned evaluation dataset; assess whether a learned model improves on the explicit rules.
- Release gates and ownership. Assign responsibility for spatial data quality, interaction behavior, accessibility and operational acceptance. Capture the baseline and candidate under the same conditions, preserve the release artifact and document a rollback path. Pilot acceptance should require accurate total counts, discoverable critical exceptions, recoverable query errors and accessible alternatives to drawing. For a later evaluation dataset, retain scenario provenance and reviewer evidence so a training label can be audited. A rollout decision should use that evidence and any unresolved limitation, rather than the apparent polish of the candidate screen.

Appendix

A. Glossary

Term	Definition
Basemap	Geographic background beneath operational layers.

Term	Definition
Occlusion	Overlap that conceals useful information.
Cluster	Display grouping of nearby features; not a query boundary.
Spatial predicate	Rule such as intersects that determines geometric selection.
Request generation	Version used to prevent stale query results replacing newer state.
WCAG AA	Conformance level spanning more than contrast alone.

Table 5. Terms used in this document.

B. Sources and Research Notes

[SRC] Primary source: Case Studies/NON Template/Geospatial_AI_Critique_Guide.pdf. Related documents in the supplied archive inform cross-project context. Research checked 7 October 2026. Additions are proposed designs or validation methods unless explicitly attributed to source evidence.

[W1] W3C. Contrast Minimum. https://www.w3.org/WAI/WCAG22/Understanding/contrast-minimum.html

[W2] W3C. Non Text Contrast. https://www.w3.org/WAI/WCAG22/Understanding/non-text-contrast.html

[W3] W3C. What Is New in WCAG 2.2. https://www.w3.org/WAI/standards-guidelines/wcag/new-in-22/

[G1] Esri. Feature Reduction Cluster. https://developers.arcgis.com/javascript/latest/references/core/layers/support/FeatureReductionCluster/

[G2] Esri. Feature Layer Queries. https://developers.arcgis.com/javascript/latest/references/core/layers/FeatureLayer/

[A1] NIST. AI Risk Management Framework. https://www.nist.gov/itl/ai-risk-management-framework

Method and limitations. Source assertions, representative simulations, illustrative graphics, researched implementation guidance and proposed acceptance targets are kept distinct. Research can support design choices but cannot establish missing historical client facts, actual delivery dates, measured results or testimonials.

C. About the Author

Keith's life has been shaped by curiosity, adventure, service, and a willingness to step into unfamiliar territory. He is married to his best friend and is the proud father of two Auburn University graduates, both thriving in their respective careers. For all the places he has been to, and the experiences he has gathered, his family remains the center of his life.

His interests span, archaeology and biblical research, skydiving, base jumping, and rock climbing, writing and drawing, and design and development. Across those pursuits runs a common thread: a desire to explore, understand, and create. He is as drawn to questions about the ancient world as he is to the possibilities taking shape in the next generation of technology.

Keith has traveled to every continent except Antarctica, lived in the United States, India, and Israel, and spent more than 20 years immersed in international cultures. Those experiences have deepened his appreciation for different perspectives and the importance of understanding people within the context of their lives. He served in the U.S. Army as a team leader with a special operations unit under Joint Special Operations Command (JSOC) and later supported global anti-human-trafficking organizations through field-based roles.

Today, Keith brings that breadth of experience to his work in AI strategy, design, and development. He sees the present as one of the most extraordinary periods of technological change in human history—and approaches it with the same curiosity that has carried him across the world. His focus is on turning emerging possibilities into useful tools, thoughtful experiences, and businesses that solve real problems. He remains an explorer at heart, with more to learn, more to build, and more to contribute.

Keith Trowbridge

AI Product Designer and AI Strategy Consultant

SarDev AI Labs

keith@keith-trowbridge.com

linkedin.com/in/keith-trowbridge

keith-trowbridge.com