ENKIDU

— MIOVSKI GROUP —

CASE STUDY

Trust-Assured Agentic AI

How the Trust Assurance Framework and Tangram shaped a multi-agent capability-assessment system for Environment and Climate Change Canada.

English · Complete edition

Version 1.0

21 June 2026

Prepared by MIOVSKI GROUP

Classification · Unclassified

Glossary

Key terms used throughout this case study. Framework terms are drawn from the Tangram and Trust Assurance overviews; technical terms describe the deployed ECCC system.

Zone (Tangram)

A bounded operational space within a process, defined by a single goal and explicit entry and exit conditions. A zone may be physical, logical, or both.

Continuum (Tangram)

A connected set of zones that together form a complete end-to-end process. In this engagement, the Digital Roadmap pipeline is modelled as one continuum.

Policy (Tangram)

An instrument of objective governance that answers “Is this permissible?” Given the facts of a situation, a well-formed policy returns a deterministic decision.

Conops

Operational service delivery — the staff, procedures, tools, and workflows that answer “How do we do it?” within a zone.

Calibrated Trust

Reliance matched to a system’s actual, demonstrated trustworthiness. Both over-reliance and under-reliance are treated as failures.

Assurance Case

A structured argument in which a trust claim is decomposed into sub-claims and tied to concrete evidence — the Claim, Argument, Evidence structure.

Trust Harness

A runtime layer that wraps an AI-augmented system and applies the assurance framework to every live request, verifying each step against policy before allowing it to proceed.

Trust Zone

A graduated capability envelope an agent earns through demonstrated good behaviour. Higher zones grant more capability but raise the entry bar and tighten monitoring.

Provenance / Chain of Evidence

A signed, hash-linked, tamper-evident record of who or what produced each piece of evidence, how, and from what inputs — making every claim reproducible and auditable.

MCP (Model Context Protocol)

An open integration standard that exposes external capabilities — document I/O, lookups, scoring — to agents as well-typed, callable tools.

Agent

A narrowly-scoped reasoning unit with an explicit role, an output contract, and a declared set of tools it is permitted to call.

ECCC

Environment and Climate Change Canada, the federal department for which the proof of concept was developed.

Abstract

Environment and Climate Change Canada (ECCC) needed to assess the maturity of its business capabilities and to assemble branch-level Digital Roadmaps from a large, heterogeneous body of internal documentation — work that is slow, expensive, and subjective when performed by hand. Enkidu built a multi-agent, agentic AI proof of concept (PoC) to perform this synthesis faithfully, transparently, and repeatably, running entirely on local infrastructure to preserve data sovereignty.

This case study describes how two Enkidu frameworks were applied to that engagement. Tangram supplied the operational architecture — decomposing the assessment and roadmap pipeline into bounded, governed zones with explicit handoffs. The Trust Assurance Framework supplied the discipline that makes the system's outputs defensible — every trust claim bound to a policy, supported by a measurable outcome, justified by an annotated chain of evidence, and traceable through a tamper-evident provenance record. The ECCC PoC was, in effect, the first end-to-end proving ground for both frameworks.

The result is a system whose central design rule — determinism where it matters, reasoning where it helps — is a direct expression of the Tangram policy-purity principle and the Trust Assurance separation of duties. Scoring is code-enforced ground truth; the language model interprets, extracts evidence, and drafts narrative, but never assigns a final score. This case study shows, component by component, how the frameworks shaped that outcome.

Don't ask to be trusted — show, claim by claim and step by step, exactly why you can be.

1. Introduction

Organizations routinely need to understand the maturity of their business capabilities in order to prioritize investment, drive transformation, and manage risk. A great deal of the relevant evidence already exists inside the organization — interview transcripts, process documents, incident post-mortems, change logs, audit findings — but reading and synthesizing it at scale is impractical for humans, and traditional consultant-led assessments are expensive, subjective, and difficult to refresh.

ECCC engaged Enkidu to demonstrate that a well-constrained agentic AI system could perform this synthesis at scale while remaining trustworthy enough to inform real decisions. The deliverables span a capability-maturity assessment and a set of branch-level Digital Roadmaps, each assembled from authoritative departmental sources with full citation traceability. Because those outputs feed investment and transformation choices, the bar was not merely “useful” but “defensible” — an analyst, an auditor, or a senior executive must be able to ask of any number in the report, “where did this come from, and why should I believe it?” and receive a complete answer.

That requirement is exactly what the Trust Assurance Framework is built to satisfy, and the operational complexity of a multi-stage, multi-agent pipeline is exactly what Tangram is built to make legible. This engagement therefore did double duty: it delivered a working system for ECCC, and it served as the first real proving ground for both frameworks together. The remainder of this document explains the two frameworks briefly, then shows how each one shaped the ECCC solution in practice.

What this case study shows

How Tangram structured the assessment-and-roadmap pipeline into bounded, governed zones with explicit, contractual handoffs.

How the Trust Assurance Framework turned “we believe this system is trustworthy” into a structured, evidenced position — at both release time and inference time.

How the two frameworks reinforced one another, and how the PoC was extended into a full Digital Roadmap continuum.

2. The Frameworks Applied

Two Enkidu frameworks were applied to the ECCC engagement. They operate at different levels — Tangram governs the shape of the operation, Trust Assurance governs the trustworthiness of what flows through it — and they are deliberately complementary. This section summarizes each before the case study shows them at work.

2.1 Tangram — Operational Architecture

Tangram is a business-architecture framework whose central insight is that every operation can be decomposed into discrete, bounded service zones, each with explicit entry conditions, exit conditions, a governing policy layer, and an operational delivery layer. By making those boundaries legible and measurable, Tangram turns vague operational complexity into a navigable, auditable, improvable system.

Three Tangram ideas matter most for this engagement. First, the zone: a unit of work with one clear goal and two boundary conditions, where the exit condition of one zone becomes the entry stake of the next, creating a chain of contractual handoffs. Second, policy purity: a policy answers only “is this permissible?”, never “how do we do it?” — keeping governance separate from operations so that policy effectiveness can be measured independently. Third, the error zone: failures are first-class citizens with their own entry conditions, operations, and a structured route back into the main flow, rather than ad-hoc exceptions.

You cannot manage what you cannot measure, and you cannot measure what you have not defined. Tangram makes your operations definable.

2.2 Trust Assurance — Calibrated, Evidenced Trust

The Trust Assurance Framework is a policy-based, evidence-backed, provenance-tracked approach to asserting and measuring trust in AI-augmented systems. Its goal is not to maximize trust but to produce calibrated trust — reliance matched to the system’s actual, demonstrated trustworthiness — because both over-reliance and under-reliance are failures.

Every trust claim is expressed as an assurance case: a claim, the argument that decomposes it, and the concrete evidence that supports it. Each claim is bound to exactly one policy with an explicit acceptance threshold, and is satisfied only when two conditions both hold — the measured outcome meets the threshold, and the supporting evidence chain validates. The framework assures nine trust components, each mapped to the NIST AI Risk Management Framework, and records everything as signed, hash-linked provenance so that every claim is reproducible and auditable.

The same machinery runs at inference time in the Trust Harness: a zero-trust reference monitor that verifies every step an agent takes before allowing it to proceed, with graduated trust zones that grant capability only as it is earned and a strict separation of duties between the components that assess, judge, and act. The framework is aligned throughout to recognized standards — NIST AI RMF, ISO/IEC 42001, NIST SP 800-207 zero-trust, and W3C PROV — so that the resulting trust dossier is externally defensible.

- Framework — What it governs — Contribution to the ECCC PoC
- Tangram — The shape of the operation — Decomposed the assessment and roadmap pipeline into bounded zones with contractual handoffs, pure policies, and designed error paths.
- Trust Assurance — The trustworthiness of the work — Bound every score and claim to a policy, an evidence chain, and signed provenance; enforced zero-trust verification and earned-capability zones at runtime.

Table 1 · The two frameworks operate at different levels and reinforce one another.

3. The ECCC Solution: A Multi-Agent System

The PoC is a multi-agent system that ingests a heterogeneous body of internal documentation — interview transcripts, process documents, incident reports, change-request summaries, audit findings — applies anchored, deterministic rubrics, and produces criticality scores, maturity scores, and a structured narrative of pain points, challenges, and opportunities mapped to a pre-defined capability model. Every output carries its supporting evidence.

The entire pipeline runs on local infrastructure to preserve data confidentiality: an NVIDIA DGX Spark workstation hosts the reasoning model (Llama-3.3-Nemotron-Super-49B-v1.5, served through an NVIDIA NIM microservice), a Chroma vector store for document memory, a lightweight relational store for the capability model and run history, Arize Phoenix for observability, and a set of Model Context Protocol (MCP) servers for document I/O and tool integration. Nothing leaves the workstation; the system is sealed from the public internet at the tool layer by design.

3.1 The Central Design Rule

The architecture’s foundational premise is determinism where it matters, reasoning where it helps. Scoring formulas, rubrics, weighting logic, and the capability model are treated as authoritative, code-enforced ground truth. The language model is used for interpretation, evidence extraction, and narrative synthesis — never for arithmetic or final-score assignment. This single rule is the seam where both frameworks attach: it is the Tangram policy-purity principle (governance separated from operations) and the Trust Assurance separation of duties (the component that judges is not the component that acts), expressed as one engineering decision.

The system never produces “just” a number. It produces a number plus its full, replayable derivation.

3.2 Agents and Tools

A supervisor/worker orchestration pattern is used rather than a free-form swarm: a supervisor plans the run and dispatches work to specialist agents, each implemented as a workflow node with a narrow role, an explicit output contract, and a declared set of permitted tools. A deliberate design rule keeps the agent that extracts evidence separate from the agent that assigns a score, which is separate again from the agent that writes the narrative — mirroring separation of duties and substantially reducing confirmation bias and hallucination.

- Agent — Responsibility
- Supervisor — Plans the assessment run, selects capabilities to analyze, dispatches work, and enforces completion criteria and cost caps.
- Retrieval — Retrieves the most relevant document chunks from Chroma using vector similarity plus metadata filters (type, weight, date).
- Evidence Extraction — Reads retrieved chunks and extracts structured evidence items — pain points, challenges, opportunities — each with a source citation.
- Scoring — Applies the anchored rubric to the evidence and calls the Scoring MCP to compute the final weighted value in code.
- Critique — Independently reviews the scoring output against the evidence and rubric, flags inconsistencies, and can force a rescore.
- Aggregator — Consolidates per-capability results into the organization-level view and identifies systemic patterns.
- Narrative — Produces the human-readable report narrative, writing nothing that is not grounded in structured scores and cited evidence.

Table 2 · The fixed set of specialist agent roles, each narrowly scoped and separately permissioned.

Tools are the primary control point for correctness and safety. Custom MCP servers expose the deterministic rubric lookups, the capability taxonomy, and the report assembly as well-typed, callable tools, so that scoring logic lives in code rather than in a prompt. Every tool invocation — arguments, return value, duration, and the issuing agent — is captured as a Phoenix span, which is also the primary mechanism for detecting prompt-injection attempts.

4. Tangram in Practice: The Roadmap Continuum

The ECCC pipeline maps cleanly onto Tangram. The end-to-end flow — from raw branch documents to a completed Digital Roadmap — is a single continuum, and each analytical stage is a zone with one goal and explicit boundary conditions. The exit condition of each zone (a validated, cited artifact) becomes the entry stake of the next, so the handoffs are contracts rather than implicit assumptions.

This is where the prompts that drive the system become visible as zone definitions. Each analysis prompt specifies a role, authoritative sources (a hard constraint), a mandatory phased workflow, citation requirements, and a coverage gate — in Tangram terms, a zone goal, its entry stakes, its conops, and its exit conditions. The BP Analysis stage, for example, opens with a Phase 0 source-confirmation step that halts if the methodological reference document is missing: a textbook entry condition that refuses to begin work until its contract is satisfied.

- Zone (pipeline stage) — Zone goal — Exit condition / entry stake for the next zone
- BP Analysis — Turn a Branch Business Plan into strategic findings, ranked initiatives, and an atomic action-level matrix. — A validated workbook that passes the coverage gate — every finding and initiative represented, no missing critical values.
- Objectives & Drivers — Derive the roadmap's strategic anchor: business objectives, drivers, and sponsors. — A populated Section 2, every cell traceable to the BP Analysis workbook.
- Current State — Baseline the As-Is across capabilities, applications, data, and pain points. — An evidence-based current-state view scored against maturity dimensions.
- Target State — Derive the To-Be vision from the departmental plan, filtered to the branch. — A target-state vision from which every later initiative and milestone derives.
- Initiatives — List the discrete initiatives that move the branch from current to target state. — An initiative portfolio with owners, deliverables, and metrics — all cited.
- Technology Recommendations — Produce the technology-branch response, anchored to objectives and initiatives. — Technology recommendations, each anchored to an objective and an initiative.

Table 3 · The Digital Roadmap continuum, expressed as Tangram zones with contractual handoffs.

4.1 Policy Purity and Coverage Gates

Tangram's policy-purity principle — that a policy answers only "is this permissible?" — appears directly in the pipeline's hard constraints. The rule "do not use outside knowledge; if sources are insufficient, say so; if sources disagree, surface the conflict" is a pure policy: given the facts of a cell, it returns a deterministic verdict (cite it, flag it as not specified, or surface a conflict) without prescribing how the analysis is performed. Keeping that rule separate from the operational work is what makes compliance measurable — the coverage gate can then test, mechanically, whether every finding and initiative is represented and whether any critical value is missing.

4.2 Errors as First-Class Zones

Each stage also designs its own error path rather than deferring failure handling. A coverage-gate failure does not silently drop items; it produces a Coverage Gate Report that lists each gap, attempts to find supporting evidence, and only marks an item uncovered when no evidence exists. A branch-name mismatch halts the run and asks the user to confirm scope. These are Tangram error zones: defined, measurable, and with a structured route back into the main flow.

If the only error handling in your design is "contact a supervisor," you have an undesigned zone. The ECCC pipeline designs its failures in advance.

5. Trust Assurance in Practice

Where Tangram shapes the pipeline, the Trust Assurance Framework makes its outputs defensible. The PoC was the first system to exercise the framework end-to-end, and the mapping is concrete: the framework's nine trust components, its provenance model, and its inference-time Trust Harness each correspond to a specific feature of the deployed system.

5.1 The Nine Components, Realized

Each of the framework's nine trust components is realized by a deliberate architectural choice rather than by assertion. The table below maps the components to the mechanisms that deliver them in the ECCC system.

- Trust Component — How the ECCC system realizes it
- Competence & Reliability — Anchored, deterministic rubrics and a fixed capability model; performance gated against a curated gold-standard evaluation dataset.
- Transparency & Explainability — Structured evidence-then-judgment reasoning; every score traces to cited evidence and an applied rubric band.
- Calibrated Confidence — Confident-wrongness guarded by an independent critique agent and groundedness evaluators; low-support items are dropped, not asserted.
- Predictability & Consistency — Fixed model version, versioned prompts, and deterministic scoring formulas make runs reproducible and trendable over time.
- Accountability & Recourse — Full audit trail of documents, rubrics, tool calls, and scores; a human-in-the-loop checkpoint before any report is released.
- Fairness & Bias Management — Disaggregated maturity scoring per directorate and dimension, surfacing weak subgroups rather than letting an average mask them.
- Security & Robustness — Document sanitization, prompt-injection input rails, per-agent tool allow-lists, and no network egress at the tool layer.
- Alignment with Intent — Scope enforcement keeps agents within the capability-assessment remit; the model never assigns final scores.
- Human Agency & Reliance — The PoC is explicitly decision-support, not an autonomous decision-maker; the final report is reviewed by a human analyst.

Table 4 · The nine trust components, each mapped to a concrete mechanism in the deployed system.

5.2 Provenance and the Chain of Evidence

The framework's insistence on a tamper-evident chain of evidence is realized through structured output contracts, citation enforcement, and full observability. Every agent-to-agent message uses a validated schema; free-form text is confined to fields explicitly labelled as narrative. Every evidence item and score must carry a source citation resolvable to a specific document and chunk, and outputs lacking citations are rejected. Every run is persisted with a unique identifier — which documents were considered, which rubrics applied, which scores computed — so that the question "how has this capability's maturity changed since the last run?" is answerable, and so that any historical result can be reproduced or explained.

The difference the framework insists on: not "a number someone typed," but "a number a named, verifiable process produced."

5.3 The Trust Harness at Inference Time

The framework's most distinctive contribution is moving claims, policies, evidence, and provenance to inference time. The ECCC system embodies the Trust Harness as a zero-trust broker over agent behaviour:

nothing an agent produces — plans, inter-agent messages, tool calls, or outputs — is trusted by default. The three enforcement phases map directly onto the deployed guardrails.

- Harness phase — What it checks in the ECCC system
- Pre-flight (ingress) — Document sanitization strips active content; input rails scan for injection patterns and reject out-of-scope operator requests before anything enters an agent context.
- In-flight (per step) — Per-agent tool allow-lists enforce least privilege; the Filesystem MCP is bounded to designated input/output directories; per-run tool budgets prevent runaway loops.
- Post-flight (egress) — Schema validation, citation enforcement, groundedness checks, and PII screening run before any narrative is written to the final report.

Table 5 · The Trust Harness’s three enforcement phases, realized as the PoC’s layered guardrails.

Two further harness principles are visible in the design. Security and high-stakes checks fail closed — an output that fails validation is rejected and retried within a fixed budget rather than allowed through. And no single probabilistic check is ever a sole gate: the deterministic Scoring MCP, the independent critique agent, and citation enforcement bound the impact if any one check is wrong. This is the framework’s defence-in-depth, instantiated.

5.4 Earned Capability and Separation of Duties

The framework’s graduated trust zones — capability earned through demonstrated good behaviour — appear in the PoC’s permissioning model: most tools are read-only, only the report-builder and a bounded filesystem writer are mutating, and write access is the most tightly scoped capability in the system, granted to the fewest agents. The framework’s separation of duties — that no component can both judge that an agent deserves more power and grant it — is mirrored in the agent topology: the evidence extractor cannot call the report builder, and the narrative agent cannot call the Scoring MCP to change a score after the fact. The question “who granted this agent the power it used?” is always answerable from the trace.

6. Extending the Proof of Concept

With the trust-assured PoC established, the engagement extended it from a single capability-assessment stage into the full Digital Roadmap continuum described in Section 4. Crucially, each extension was added the same way: a new zone with explicit entry stakes, a pure source-and-citation policy, a designed coverage-gate error path, and the same provenance discipline. New functionality did not relax the trust posture — it inherited it.

The extensions added new analytical zones (Sections 2 through 6 of the roadmap), each driven by a versioned prompt that adapts the original BP Analysis methodology to a new deliverable and a new evidentiary source. Several extensions also introduced new ingestion inputs — a branch maturity assessment scored across twenty-three capability dimensions, a deterministically-scored pain-point analysis report with inter-rater reliability statistics, and a pain-point-to-capability impact mapping — each normalized through the ingestion pipeline before it could be cited. The pattern is that every new MCP, agent, or input is admitted only once it can participate in the chain of evidence.

How the frameworks made extension safe

Tangram: each new stage is a self-contained zone, so it could be added without destabilizing the others; its handoff contract is explicit and its error path is designed in advance.

Trust Assurance: each new source is brought under the same citation, provenance, and coverage discipline, so the extended system is no less defensible than the original PoC.

Both together: version-pinned prompts and a fixed model mean every extended result remains reproducible and auditable, exactly as the original assessment was.

Introducing new capability into a governed zone does not change the zone’s governance requirements — it intensifies them.

7. Outcomes & Conclusion

The ECCC engagement delivered a working, on-premise, multi-agent system that assesses business-capability maturity and assembles branch Digital Roadmaps from authoritative sources — faithfully, transparently, and repeatably. Just as importantly, it proved both Enkidu frameworks in a single real deployment.

Tangram gave the engagement an operational architecture that an enterprise architect can read at a glance: a continuum of bounded zones, contractual handoffs, pure policies, and designed error paths. The Trust Assurance Framework gave the same engagement a defensible trust position: nine components each realized by a concrete mechanism, a tamper-evident chain of evidence behind every score, and a zero-trust harness that verifies every step at inference time. The two are not separate overlays; they meet in the system's central rule — determinism where it matters, reasoning where it helps — which is at once a Tangram policy boundary and a Trust Assurance separation of duties.

The outcome is the framework's stated objective made real: not maximal trust, but calibrated trust — reliance matched to demonstrated trustworthiness — backed by a verifiable dossier that an analyst, an auditor, or a senior ECCC executive can interrogate. The PoC is complete; the frameworks that shaped it are now proven, and the pattern it establishes is reusable for the next governed, AI-augmented deployment.

Calibrated trust, earned claim by claim and step by step — and an operational architecture legible enough to prove it.

References

[1] Enkidu Enterprises (2026). Tangram Overview: A Structured Framework for Operational Architecture, Governance, and AI-Augmented Service Delivery. Version 1.0.

[2] Enkidu Enterprises (2026). Trust Assurance Overview: A Policy-Based, Evidence-Backed Approach to Asserting and Measuring Trust in AI-Augmented Systems. Version 1.0.

[3] Enkidu Enterprises (2026). Agentic AI Proof of Concept — Business Capability Maturity Assessment: Architecture and Design Document.

[4] National Institute of Standards and Technology (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0).

[5] ISO/IEC 42001:2023. Information technology — Artificial intelligence — Management system.

[6] National Institute of Standards and Technology (2020). SP 800-207, Zero Trust Architecture.

[7] W3C (2013). PROV-DM: The PROV Data Model.