ARIZONA WESTERN COLLEGE • MATADOR WELLNESS CENTER •
ARCHITECTURE BRIEF 01

Sovereign Clinical AI, DOM GraphRAG & W3C PROV-O Provenance

Deterministic & Agentic Curriculum-as-Code Engineering Across the 7-Course EXW Fleet (S-A-R Standard v2.2)

S–A–R v2.2 • WCAG 2.2 AA (9:1)

In Plain English (Executive Summary):
Before you read the technical details below, here is the short version: **The EXW Clinical Auditor is a 100% free, local, privacy-preserving clinical verification engine built for Arizona Western College Exercise Science & Wellness students.** Unlike generic cloud chatbots that guess from unverified web text, our system combines a **3,151-chunk DOM Knowledge Graph (DOM GraphRAG)**, **100% local Apple Silicon inference (Ollama)**, and **W3C PROV-O provenance receipts** so students act as the *Clinical Architect/Pilot* and AI serves strictly as an auditable *Clinical Calculator*.

1. Architectural Problem Statement: Transcending Flat “Docs-as-Code”

Traditional “Docs-as-Code” pipelines treat Markdown files as isolated, flat text blobs in Git. When a standard vector Retrieval-Augmented Generation (RAG) pipeline chunks flat Markdown files without document-object-model (DOM) hierarchy or domain ontology edges, it suffers from context fragmentation and cross-course guideline drift (`SSOT_LEAK`).

In the **EXW Master Library (`S–A–R v2.2` : Scalable / Auditable / Repeatable)**, every curriculum module is governed as a **structured semantic graph** connecting the Course, Module, DOM Heading Path (`h2 > h3`), Course Learning Outcome (`CLO1–CLO3`), Clinical Framing Role (`Pattern / Rule / Solve`), and Federal Single Source of Truth (SSoT).

2. Layer 1: DOM GraphRAG & Course-Scoped Vector Retrieval (3,151 Chunks)

Implemented in `scripts/index_corpus.py` and `scripts/query_corpus_rag.py` , each of the **3,151 indexed curriculum chunks** carries both a 768-dimensional `nomic-embed-text` vector (cached by SHA-256 content hash) and an explicit `dom_graph_node` payload:

DOM GraphRAG Field	Engineering Specification	Example Value (EXW 101 Module 05)
`dom_path`	Hierarchical DOM heading lineage (`h2 > h3`)	`h2: Assigned Borderland Case Data (Yuma, AZ / Imperial Valley, CA)`
`clinical_framing_role`	Deterministic classification into the 6 pedagogical roles	`Clinical_Case_Pattern` , `SSoT_Rule_Constraint` , `Instructional_Core_Solve` , `Active_Verification` , `Exemplar_10_10` , `Plain_English_Callout`

DOM GraphRAG Field	Engineering Specification	Example Value (EXW 101 Module 05)
`clo_alignment`	Explicit mapping to institutional Course Learning Outcomes	`EXW101-CLO1`, `EXW101-CLO2 (PAGA 2018)`, `EXW101-CLO3`
`graph_edges`	RDF-style `[subject, predicate, object]` triples	`partOfModule`, `governedBySSoT`, `fulfillsClinicalRole`, `alignsWithCLO`

3. Layer 2: Sovereign Apple Silicon Inference & Bilingual Norteño Parity

- **Zero Cloud Egress (FERPA Sovereignty):** All clinical audits execute 100% locally on Apple Silicon M1/M2 Unified Memory via custom Ollama Modelfiles (`exw-argon4-cac:latest` and `exw-gemma4:latest`) served by a multi-worker **Gunicorn + Flask WSGI** backend (`127.0.0.1:5050`).
- **The Centaur Protocol & Active Verification Gate:** Students encapsulate case vitals inside `###CASE_DATA_START### ... ###CASE_DATA_END###` and author a mandatory **150-word Active Verification ("Why" Check)** in their own voice before receiving a **10 / 7 / 0 Mastery Scale** pre-submission audit.
- **Hispanic-Serving Institution (HSI) Bilingual Parity:** Submissions authored in English receive feedback opening with `In plain English:`; submissions authored in Spanish or regional Norteño-Spanish receive authentic Norteño-Spanish coaching opening with `Hablando claro:` or `En palabras sencillas:`.

4. Layer 3: W3C PROV-O Provenance Receipts (`http://www.w3.org/ns/prov#`)

Every audit response returned by `POST /api/v1/audit` and `POST /audit/submission` attaches a machine-verifiable **W3C PROV-O Provenance Receipt** linking the output to the exact model, governing SSoT authority, and retrieved DOM GraphRAG nodes:


```
{
  "@context": "http://www.w3.org/ns/prov#",
  "prov_type": "prov:Entity",
  "audit_receipt_id": "EXW-PROV-EXW101-83A9BD8DDBC4",
  "generatedAtTime": "2026-11-01T17:10:09Z",
  "wasGeneratedBy": {
    "activity": "EXW_Clinical_Auditor_S-A-R_v2.2",
    "framework": "Clinical Framing Protocol (Pattern / Rule / Solve)",
    "active_verification_gate": "150-Word Verbatim Defense Check (10-7-0 Scale)"
  },
  "wasAttributedTo": {
    "agent": "exw-argon4-cac:latest",
    "architect": "Coach Alan (Arizona Western College — Matador Wellness Center)",
    "ferpa_sovereignty": "100% Local Apple Silicon Inference (Zero Cloud Egress)"
  },
  "wasDerivedFrom": {
    "ssot_authority": "Physical Activity Guidelines for Americans (PAGA), 2nd Edition (2018)",
    "retrieved_chunk_sources": ["EXW101:Module 05:module-06_public_health.md#Assigned Borderland
Case Data"]
```

```
  }
}
```

5. Layer 4: WCAG 2.2 AA (9:1 Double-Contrast Guard) & 10/10 Accessibility Discipline

All 14 live Canvas LMS deployment targets across all 7 courses enforce a minimum **9:1 luminance contrast ratio** (exceeding WCAG 2.2 AA's 4.5:1 baseline), semantic HTML5 landmarks (`<pre><code>`, native `<button>`, `<label for>`), full keyboard focus trapping (`Tab` / `Shift+Tab` / `Escape`), and 4-stage `aria-live="polite"` telemetry.

Author: Alan Pruitt ("Coach Alan"), Exercise Science & Wellness Faculty, Arizona Western College

License: CC BY-NC-SA 4.0 • alanpruitt.com • exwdemo.dev