

## 01 THE CURRENT PROBLEM

Rules provide the backbone of benefits eligibility and enrollment systems. Today they are **sourced** from an array of policy documents, manuals, emails, or sometimes no documented source, then interpreted into closed systems out of public view — making them hard to verify and harder to fix. Yet a top ask from state leaders rebuilding their systems is **traceability and transparency** in their rules engines.

*"Right now, the determinations come back, and it's hard to know exactly what happened... Our caseworkers tell us that they spend a lot of time trying to figure out what led to that, and where they make a data entry error that is causing the determination to come out differently than what they were expecting. So having that traceability is also really huge." – State Leader (source)*

Current systems create operational risk, slow error correction, and deepen vendor entrenchment: when the only operational version of a rule lives inside a closed system, governments lose leverage to inspect, compare, reuse, or change it. The result is concentrated vendor lock-in, expensive one-off change requests, and large budget exposure when eligibility accuracy breaks down — especially acute as states implement **H.R. 1 (OBBBA)**, with substantial rule changes and cost implications.

### \$6B+

in Deloitte-run eligibility system contracts across 25 states

KFF Health News found Deloitte alone holds at least \$6 billion in state eligibility system contracts — showing how concentrated and entrenched the current vendor market is. [KFF Health News](#)

### \$1.6M – \$20M

for recent state change requests tied to one federal policy wave

Kentucky, Vermont, Illinois, and Iowa estimates range from \$1.6M to \$20M just to implement new OBBBA Medicaid changes. [KFF Health News](#)

### 27 states

could face \$100M+ annual SNAP cost shifts tied to error-rate formulas

CBPP estimates 27 states would face projected SNAP cost shifts above \$100M per year, on top of the law's reduced federal share for administration. [CBPP](#)

This fits PBI's backend-operations and modern-infrastructure priorities because traceable, cited rules are not a standalone screener or awareness tool — they are infrastructure for safer AI, lower error rates, modular modernization, reduced vendor lock-in, and stronger government stewardship of benefits technology.

## 02 HOW WE'RE ADDRESSING THE PROBLEM

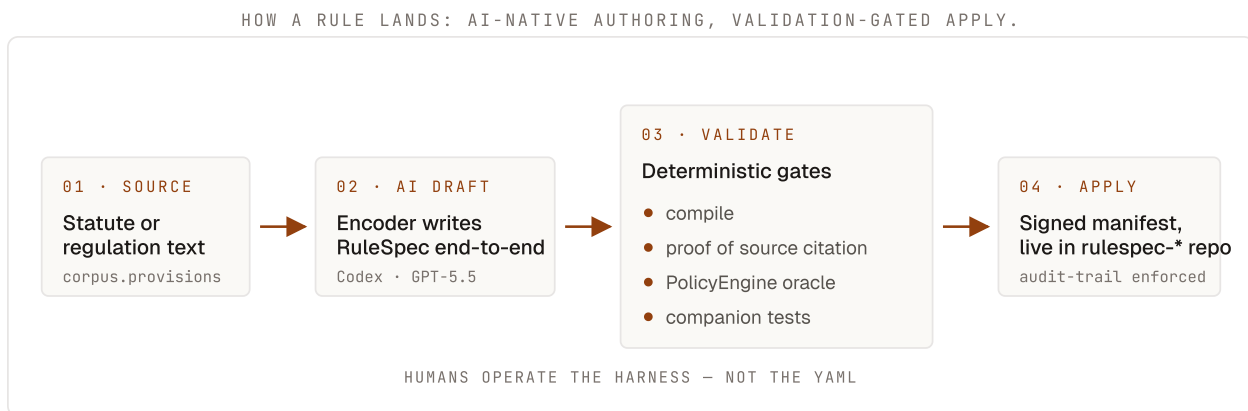
**Axiom Foundation** provides statutes, regulations, and policy rules turned into open and freely available machine-readable encodings — *cited, time-aware, executable* — so anyone can run, audit, or reform them. The encodings can be used by any tool — government systems, benefits screeners and navigation tools, cliff calculators, policy simulators, AI models — and know they are using verified, standardized rules. Axiom eliminates duplicative work and guessing at the interpretation of policy.

Three running prototypes demonstrate the system end-to-end:

- **Axiom App** — inspect source text, citations, provenance, and RuleSpec encodings for encoded rules.
- **Builder** — assemble tools powered by Axiom inputs, outputs, and explanations backed by a computation graph; integrate or download as YAML or JSON.
- **Chatbot** — answer benefit questions from an LLM constrained to encoded rules and citations.

These tools build on six years of work at **PolicyEngine** to provide infrastructure for tax and benefit analysis. PolicyEngine powers benefits navigation tools including **MyFriendBen**, **Amplifi**, and **Mirza**, and is used inside and outside of government for **policy modeling** by the Brookings Institution, the Joint Economic Committee of Congress, the Bureau of Economic Analysis, and the UK Prime Minister's office.

Axiom wouldn't have been possible even a few years ago. AI-native encoding has collapsed the cost and timeline of turning dense law into machine-readable rules: in 2023, encoding one state's tax code took roughly three FTE-months of expert human work; today our pipeline can produce a validated, citation-linked rule from a single statute or regulation section in hours, with humans operating the validation harness rather than authoring rules by hand.



### 03 WHAT'S NEEDED NOW

We have strong prototypes that demonstrate the vision and technical approach; what we need now is a pilot with government teams. Three workstreams structure the cohort.

**Validate rules and the encoding approach.** *Learning goals:* feedback and iteration on encoding approach and structure, including how traceability of rules is presented; alignment to existing systems, and how much is needed to close the gap between publicly available policy documentation and internal documentation and systems; identification of conflicts between public documentation and internal policy and code. *Impact metrics:* rule match %; % of validated test cases from various sources (unit tests, Golden, SNAP QC, CPS).

**Confirm architecture for system integration.** *Learning goals:* understand current architectures; further prototypes for integration with systems; advance approaches for presentation of rules provenance and traceability of calculations in case management and other backend systems. *Impact metric:* execution runtime against existing system.

**Design governance processes that serve public and government needs.** *Learning goals:* document and understand existing governance strategies; prototype multi-government governance; prototype how to draft, review, and resolve rule changes; test the feasibility of an AI-driven approach where governments clarify policy in source documents to be encoded by Axiom; validate the fully open and public rules approach. *Impact metric:* time, resourcing, and steps to implement a rule change or fix a discovered issue.

#### 04 PROGRAM DESIGN

PBIF support would enable us to engage multidisciplinary teams in three governments over 12 months, focused on selected SNAP, Medicaid, and WIC rule families where partner materials and validation cases are available. The pilot would test whether AI-generated, citation-backed rules encoding can help governments validate and operationalize benefit rules more quickly, transparently, and safely than current document- and vendor-mediated workflows. The hypothesis: executable rules linked to authoritative sources, tests, and discrepancy reports will improve a government's ability to identify errors, govern policy changes, and evaluate integration paths. If we are right, partner reviewers will trace determinations without reading code, validation agreement will improve or remaining discrepancies will be explainable, and each partner will make a concrete go / no-go decision about future integration.

GOVERNMENTS PROVIDE	AXIOM DELIVERS
<ul style="list-style-type: none"><li>• A multidisciplinary team of 3-5 people</li><li>• Policy documentation (guides, staff manuals, directives, internal change notes)</li><li>• Existing legacy rule logic or code, where accessible</li><li>• Test cases or validation sets, where available</li><li>• Current technical architecture and any future-state plans</li><li>• Participation in 1:1 and group working sessions</li><li>• Permission to share work processes, trainings, and shadow workers as appropriate</li><li>• Asynchronous feedback</li></ul>	<ul style="list-style-type: none"><li>• Rules encodings for selected rule families based on public sources and partner-provided documentation, guidance, cases, and code where available</li><li>• Monthly cohort meeting; recurring 1:1 working sessions per government</li><li>• In-person multi-day session in each jurisdiction including process shadowing</li><li>• Virtual learning day for other governments and field partners</li><li>• Final deliverables: open public documentation and rules, a bounded Builder-powered prototype, a grounded assistant interface using encoded rules, plus training and handoff for the government team</li><li>• Alongside the cohort, Axiom continues broader automated encoding toward PolicyEngine parity for SNAP, WIC, and Medicaid across states; that roadmap is separate from the pilot's selected-rule-family commitments</li></ul>

#### 05 TIMELINE

**Readiness period (before grant start).** Continue ingesting state statutes and regulations, improving source normalization, documenting RuleSpec, hardening the Encoder workflow, expanding validation harnesses, and improving the app and ops dashboard surfaces.

**Months 1-2 — Technical build and partner onboarding.** Gather documentation and code from each government, establish a light baseline (current traceability, available validation cases, known error categories, policy-change workflow), and begin encodings using provided documentation. By end of Month 2, each workstream should have a first rule package, companion tests, provenance records, and a partner-facing trace report.

**Months 3-4 — Deep dives and initial encoding feedback.** Multi-day working session in each jurisdiction. Compare encoded outputs against PolicyEngine where applicable, partner cases, screener outputs, manually reviewed edge cases, and any available legacy-system behavior. Classify discrepancies as source ambiguity, encoding error, implementation-policy difference, input-data gap, legacy-system behavior, or documentation conflict.

**Months 5-6 — Partner review and workflow testing.** Partner reviewers work through trace reports and discrepancy logs. Axiom observes training, lower-environment processing, and policy-change workflows where feasible. Each workstream gets one bounded prototype demonstrating a concrete use case (validating a rule change, explaining a determination, identifying a conflict between manual and public guidance, or supporting an internal eligibility review), plus an integration-options memo.

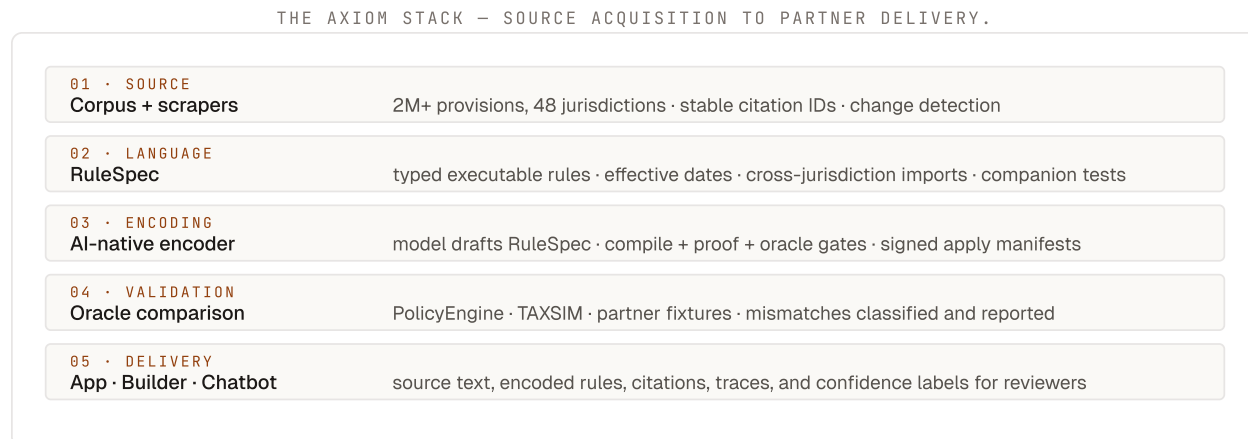
**Month 6 — Six-month proof and go / no-go.** Rerun validation, compare to baseline, and produce partner-specific recommendations on whether each workstream is ready for deeper integration, continued R&D, or a narrower deployment. Six-month proof package: validated rule packages, discrepancy taxonomies, integration architecture recommendations, and an initial governance model.

**Months 7-9 — Adoption support and governance refinement.** Test whether the six-month proof survives real operating conditions. Support partner review cycles, refine rule-update and release-governance processes, add targeted rule families where the first proof is strong, rerun validation after source updates.

**Months 10-12 — Public sharing and scale decisions.** Decide whether the approach is ready for a larger production pilot, a limited validation-sidecar deployment, or additional R&D. Virtual learning day. Public deliverables include an implementation playbook, responsible AI / governance template, traceability report examples, and a field-facing summary of what public benefits teams need to govern AI-assisted executable rules safely.

## 06 TECHNICAL APPROACH

Axiom's stack has five layers, each instrumented for review.



**Corpus and source acquisition.** Scrapers and ingestion adapters fetch authoritative statutes, regulations, and guidance from official upstreams where available, with change detection and freshness tracking. Every provision gets a stable ID, citation path, and source provenance queryable through the Corpus; source checksums, anchors, effective dates, and related metadata are preserved where available.

**RuleSpec.** A typed executable rules format with effective dates, source-claim references, cross-jurisdiction imports, durable legal IDs, companion tests, and explain traces. RuleSpec is a starting standard we expect to validate and improve with partners, not a claim about the final form governments should use.

**AI-native encoding.** Codex and OpenAI GPT backends, with GPT-5.5 as the current default, read source text, resolve corpus citations, draft RuleSpec end-to-end, and then pass deterministic compile and proof gates plus oracle checks where mappings exist before generated rules are applied through a signed manifest. Humans operate the harness — evals, prompts, validators, repair retries — not the YAML.

**Oracle comparison.** PolicyEngine, TAXSIM, screener outputs, and partner-provided cases run side-by-side with Axiom outputs where comparable; mismatches are classified and reported, not hidden.

**App and partner surfaces.** Source text, encoded rules, citations, tests, traces, run metadata, and selected program outputs are exposed through the **Axiom App**, **Builder**, and **Chatbot**. The pilot will standardize partner-facing discrepancy logs, reviewer notes, and confidence labels around those surfaces.

**Barriers.** Partner access to current manuals and artifacts; privacy or vendor restrictions on legacy code and cases; ambiguous policy sources; limited validation datasets; Q4 government availability; and who has authority to approve a rule interpretation. The pilot is designed to surface these early as reusable field guidance.

#### 07 RESPONSIBLE USE OF AI

AI accelerates drafting, extraction, mapping, and repair — never autonomous eligibility decisions. Accepted rules must be grounded in cited sources, compile into deterministic RuleSpec, pass validation, and be reviewable by humans. The pilot runs without production PII; validation cases are synthetic, de-identified, or partner-approved. When a source is incomplete or a rule cannot be validated, the system abstains, returns a bounded range, or flags the issue rather than silently encoding a guess. Reviewers — not the model — remain the final authority on whether an encoding is correct.

#### 08 TEAM AND PARTNERS

- **Max Ghenis** — Founder & CEO. Founded PolicyEngine; built the open rules-and-microsimulation stack behind Axiom.
- **Pavel Makarchuk** — Product Lead. Bridges encoded rules and usable products: partner workflow design, implementation detail, and the application layer people touch.
- **Ariel Kennan** — Advisor. In-government rules building and foundational field research on rules as code and integrated eligibility and enrollment systems.

*This concept note is intended to support scoping conversations with prospective government partners for the pilot cohort.*