## Abstract

## 1. Introduction

Australia’s transition from conventional demolition to building deconstruction and material salvage is increasingly framed as a practical pathway to a circular built environment; however, implementation remains constrained by a fragmented governance landscape that spans federal, state/territory, and local instruments. In practice, deconstruction is governed indirectly—through dispersed planning provisions, waste and resource-recovery policies, building regulations, procurement rules, and workplace requirements—rather than through a coherent regulatory “chain” that clearly defines responsibilities, establishes enabling pathways, and reduces uncertainty for industry. **[Add numbers to show the significance of deconstruction—e.g., construction and demolition waste volumes, landfill diversion rates, embodied-carbon implications, and the share of building-related materials in national waste/ emissions.]**

A central barrier is that policy language often emphasises generic “waste diversion” outcomes, which can be compatible with crushing, downcycling, and conventional demolition logistics, while offering limited specificity on **disassembly**, **component reuse**, and evidence-based certification pathways for reused materials. This definitional ambiguity makes it difficult to identify where regulation actively enables deconstruction, where it is silent, and where it unintentionally creates barriers (e.g., risk-averse procurement norms, unclear permitting requirements, or lack of reuse certification). **[Add a short paragraph that defines “deconstruction,” “selective demolition/soft strip,” “salvage,” and “reuse” as used in this paper; explicitly distinguish them from demolition and from generic recycling.]**

Methodologically, the Australian deconstruction governance landscape is difficult to analyse using manual review alone: the relevant content is distributed across heterogeneous document types (acts, regulations, strategies, guidance, and local policies), expressed in inconsistent terminology, and often embedded in long PDF instruments. This creates a need for a systematic, auditable, and queryable approach that can (i) scale to large corpora, (ii) preserve provenance back to original sources, and (iii) support structured comparison across jurisdictions. [Add a paragraph about the lack of studies using knowledge graphs + retrieval-augmented generation (RAG) for computational analysis of deconstruction governance/policy, and why evidence-linked outputs matter for policy credibility.]

This study develops and deploys a Graph Retrieval-Augmented Generation (GraphRAG) governance intelligence pipeline that converts unstructured Australian policy instruments into an evidence-linked knowledge graph and a queryable retrieval layer, enabling auditable, source-grounded analysis of how deconstruction and material salvage are governed across jurisdictions. Specifically, the pipeline implements a reproducible multi-stage workflow to (1) process heterogeneous government PDF documents into machine-actionable text, (2) retrieve and precision-filter deconstruction- and salvage-relevant clauses from broader waste narratives, and (3) populate an objective-aligned knowledge graph that encodes who governs what, where, and how—while preserving traceability from each extracted statement to its original source and verbatim evidence. The platform then supports graph traversal and query-driven synthesis to compare jurisdictions, test governance pathway completeness (e.g., Authority → Instrument → Requirement → Practice → Outcome), and identify gaps where barriers exist without corresponding enabling instruments.

Accordingly, the study pursues five objectives: designing and validating the end-to-end GraphRAG pipeline for regulatory corpus processing; developing and populating a compact knowledge-graph schema aligned to deconstruction governance; analysing semantic and definitional gaps in how “deconstruction” and “salvage” are recognised across policy instruments; mapping the institutional governance network and jurisdictional fragmentation across the three tiers of government; and performing automated gap analysis to classify governance deficits by severity (including practices with no institutional support). This research makes the following contributions:

* **C1 (Method):** An end-to-end GraphRAG pipeline for deconstruction governance that integrates semantic retrieval, deterministic LLM precision filtering, evidence-linked knowledge-graph extraction, and query-driven policy intelligence.
* **C2 (Data product):** A cleaned, auditable knowledge graph with structured entities and relations, and provenance metadata for each extracted claim (source file, chunk identifier; page number when available).
* **C3 (Analytics):** Graph-enabled analyses (pathway tracing, jurisdictional coverage, gap analysis, definitional precision) demonstrating how governance support and barriers cluster across deconstruction practices.
* **C4 (Reproducibility):** A schema-constrained extraction strategy and cleaning audit trail (merge audit, pruning log) supporting transparency and methodological replication.

## 2. Contextual Background

## 3. Research Methods and Design

This study developed and validated a computational governance intelligence pipeline to map and interrogate Australian deconstruction and material-salvage governance across federal, state, and local levels. The end-to-end workflow (see Figure 1) combines automated document acquisition, multi-engine text extraction, semantic retrieval, LLM-based precision filtering, knowledge-graph construction with evidence provenance, and a query layer that supports auditable, source-grounded answers. To construct a comprehensive yet auditable evidence base on Australian deconstruction and material-salvage governance, we implemented a structured, two-stream web search and acquisition workflow targeting (i) government regulatory instruments and (ii) industry guidance. Searches were executed using reproducible Google Custom Search–style workflows with LLM-assisted query expansion to reduce researcher selection bias and improve recall across heterogeneous terminology and document types. For the industry stream, discovery combined (a) LLM-assisted source identification, (b) organisation- and domain-targeted searches, and (c) a credibility screen integrating source reputation, recency, and topical relevance. For the regulatory stream, searches were stratified by jurisdiction (Commonwealth and each state/territory) and targeted authoritative portals and regulators (e.g., legislation repositories, EPAs, planning/building regulators, and WHS bodies). Retrieved candidates were deduplicated and scored using a regulatory-specific credibility rubric (authority tier, government tier, recency, and topical precision); only items exceeding the credibility threshold were retained and downloaded as PDFs. The combined non-academic retrieval yielded a corpus of 95 PDF documents (industry + regulatory), all of which were then subjected to full-text extraction and computational screening prior to downstream semantic retrieval and GraphRAG analysis. In this paper, “GraphRAG” refers to a retrieval–generation workflow in which (i) graph neighbourhoods and structured triples are retrieved via graph queries, (ii) supporting evidence chunks are retrieved via vector search, and (iii) a deterministic language model synthesises answers only from retrieved graph structures and cited evidence excerpts, returning “insufficient evidence” when support is missing. This architecture ensures that all outputs are traceable to specific source documents and evidence passages, distinguishing the platform from standard retrieval-augmented generation approaches that operate on unstructured text alone.

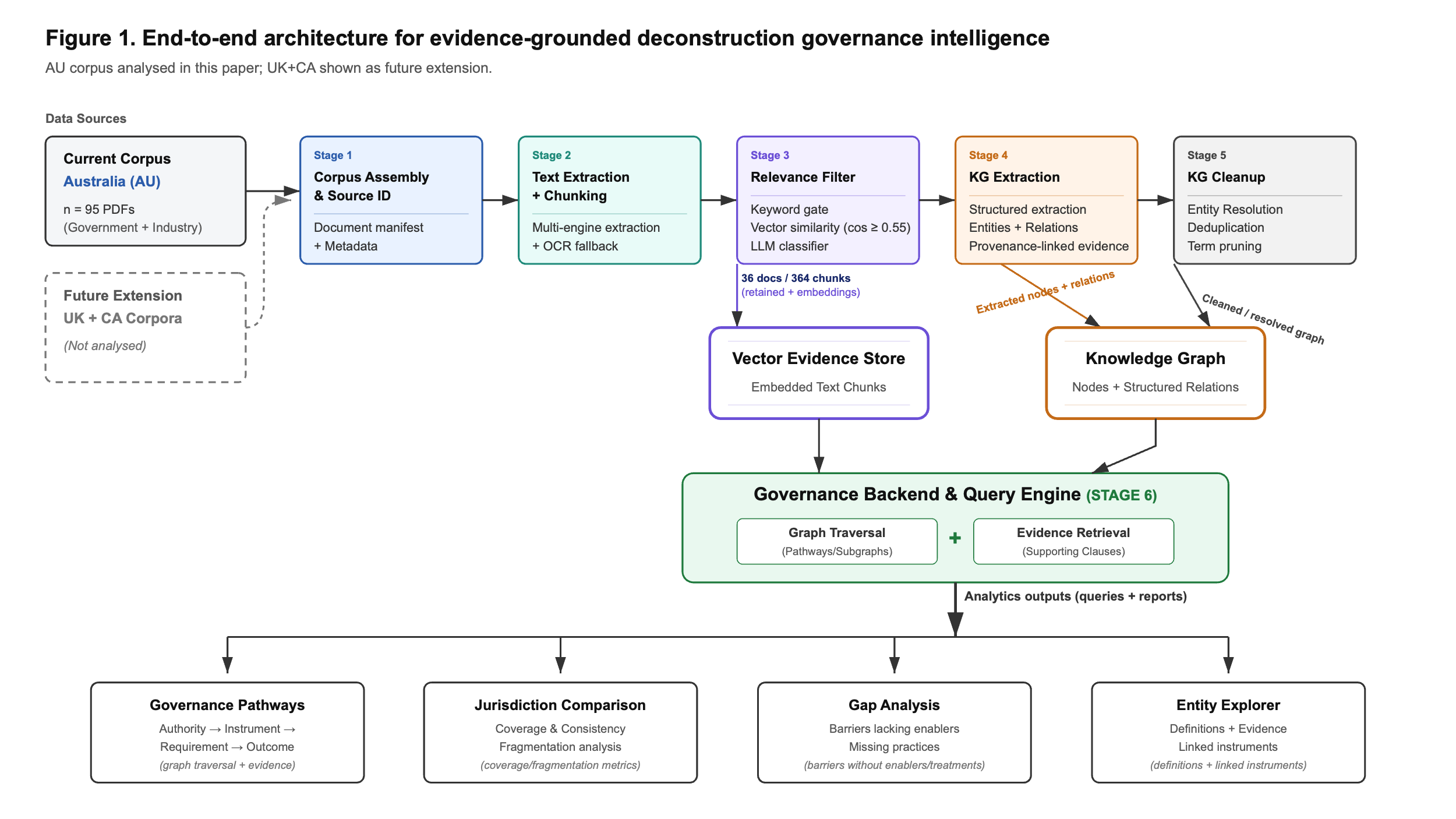


Figure 1. Evidence-grounded pipeline from PDFs to a provenance-linked knowledge graph

### 3.1 Corpus assembly and eligibility screening

To balance breadth of coverage with analytical specificity, the corpus was assembled through three complementary streams: (i) academic literature, (ii) industry guidance, and (iii) government regulatory instruments.

**Academic literature (contextual stream).** Peer-reviewed studies were retrieved via Scopus API using an LLM-assisted query expansion step to generate high-recall keyword bundles (subject, context, and Australia/jurisdiction terms). Results were deduplicated and then filtered deterministically using an explicit inclusion rubric to retain only studies materially focused on Australian deconstruction and/or material reuse. Academic outputs were retained as bibliographic metadata (e.g., DOI, title, authors) to contextualise and triangulate discussion, but were **not** processed into the knowledge graph.

**Non-academic documents (computational stream).** Industry guidance and regulatory instruments were retrieved through structured Google Custom Search workflows. The industry pipeline combined (a) LLM-assisted source discovery (to reduce researcher selection bias), (b) organisation- and domain-targeted searches, and (c) a credibility screen that integrated source reputation, recency, and topical relevance. The regulatory pipeline was stratified by Australian jurisdiction (Commonwealth, states/territories) and targeted authoritative portals and regulators (e.g., legislation sites, EPAs, planning/building regulators, WHS bodies). Retrieved documents were then scored using a regulatory-specific credibility rubric (authority tier, government tier, recency, and topical precision). Only documents exceeding the credibility threshold were retained and downloaded.

**Corpus definition used in this paper.** The combined non-academic search yielded a **retrieved corpus of 95 PDF documents** (industry + regulatory). All 95 documents were subjected to full-text extraction and computational screening. After semantic retrieval and LLM-based precision filtering, the **analysis corpus** comprised **36 documents** that contributed at least one retained chunk labelled Keep or Maybe, yielding **364 retained chunks** used for knowledge-graph construction. The remaining 59 documents were excluded because they addressed generic waste management or demolition narratives without substantive deconstruction-specific content. One document in the 36-document analysis corpus contributes only contextual (“Maybe”) passages and contains no explicit deconstruction terminology; it is included in the analysis corpus but may not appear in lexical-only counts.

### 3.2 Text extraction, segmentation, and candidate passage selection

Because the retrieved corpus included both digitally native and scanned PDFs, text extraction used a cascading, multi-engine strategy designed to maximise recall while preserving legal and tabular structure. Documents were first parsed using a layout-aware extractor to preserve structured elements (e.g., levy schedules, targets). Where extraction quality was insufficient, a robust parser was applied to resolve encoding and embedded-font issues. For scanned pages, OCR was invoked to reconstruct the text layer.

Extracted text was segmented into overlapping semantic chunks to maintain the integrity of definitions and qualifying clauses. A relatively large chunk window was used to avoid splitting legal provisions across segments. To reduce false positives and computational load, a **keyword gate** was then applied: only chunks containing high-recall deconstruction/salvage patterns (e.g., deconstruct\*, salvag\*, selective demolition, circular economy) were retained for downstream semantic retrieval.

### 3.3 Semantic retrieval and LLM-based precision filtering

Candidate chunks passing the keyword gate were embedded into a high-dimensional semantic space to enable robust matching across heterogeneous regulatory language. Semantic retrieval was implemented using a vector index with L2-distance scoring against topic-focused query vectors. A strict similarity threshold—set through pilot diagnostics—was then applied to suppress high-volume “waste diversion” noise while retaining deconstruction-specific provisions and practice-relevant clauses.

To eliminate residual ambiguity (e.g., generic recycling logistics vs. structural disassembly), a deterministic **LLM judge** was applied as a final precision layer. Each high-scoring chunk was classified at temperature 0 into one of three categories:

* **Keep:** directly addresses disassembly/deconstruction, salvage/soft strip, or explicit governance barriers/enablers for reuse
* **Maybe:** contextually relevant (institutional framing, enabling conditions) but requires cautious interpretation
* **Drop:** irrelevant to deconstruction governance (e.g., kerbside recycling, generic diversion reporting)

Chunks labelled **Keep** and **Maybe** were retained to preserve both evidence-grade clauses and enabling context. This screening step also resolves the apparent “95 vs 36” discrepancy: **95 documents were processed**, but only **36 contained retained evidence** (Keep/Maybe) and therefore contributed to the knowledge graph. The 36-document analysis corpus is defined by the GraphRAG retention criterion (≥1 Keep/Maybe chunk); lexical-only counts may yield 35 documents because one retained document contributes only contextual (“Maybe”) passages without explicit deconstruction terminology. For robustness checks, the judge label was propagated as a statement confidence weight (Keep = 1.0; Maybe = 0.6), enabling sensitivity comparisons between a “core” evidence graph (Keep only) and an expanded interpretive graph (Keep + Maybe).

### 3.4 Knowledge-graph schema and evidence-linked extraction

Knowledge-graph construction was guided by a compact, objective-aligned schema designed to represent governance mechanisms and practice pathways rather than a broad ontology. The schema comprises **ten entity types** (Instrument, Authority, Jurisdiction, Requirement, Practice, MaterialAsset, Stakeholder, Barrier, Enabler, OutcomeMetric) and **eleven relation types** that encode governance structure and causal mechanisms: ISSUED\_BY, APPLIES\_IN, APPLIES\_TO, REFERENCES, REQUIRES, PROHIBITS, INVOLVES, ENABLES, BARRIERS, AFFECTS, and PRODUCES. Relations are constrained to endpoints co-occurring within the same extracted chunk and each relation stores a verbatim evidence excerpt alongside provenance fields (source file, chunk identifier; page number when available). The schema is summarised in Figure 2. Knowledge-graph extraction was performed **at the chunk level** using a deterministic LLM configuration with enforced structured outputs. To reduce over-generation and maintain precision, extraction was constrained by (i) strict type control (entities/relations must conform to the predefined schema) and (ii) per-chunk caps on extracted entities and relations. Every extracted entity and relation was required to carry a **verbatim evidence excerpt** from the originating chunk, and each relation stored provenance metadata (source file, chunk identifier; page number when available). Relations were retained only when both endpoints were explicitly extracted from the same chunk, reducing speculative linkage.

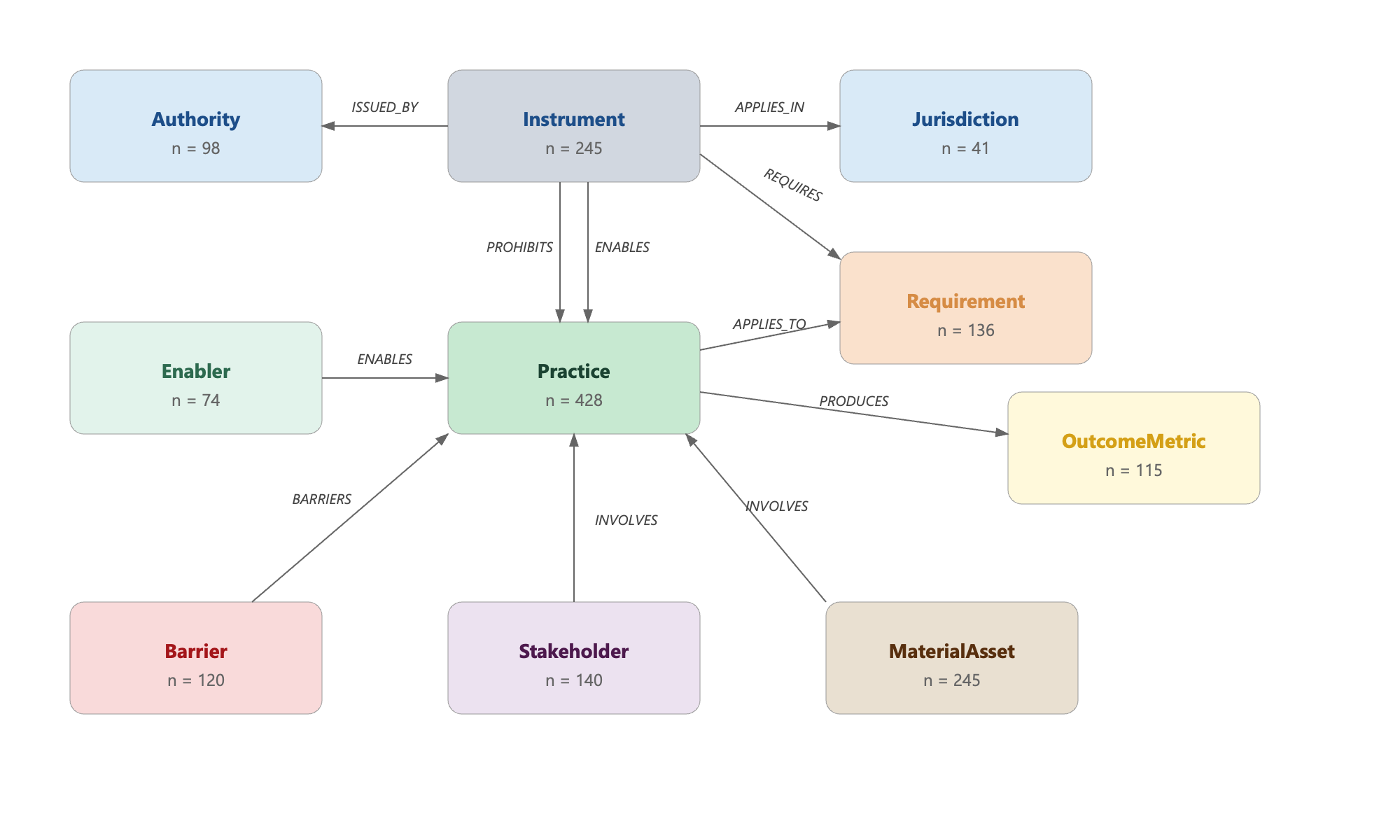


Figure 2. Knowledge Graph Schema showing the 10 entity types and 11 primary relationship types

### 3.5 Canonicalisation, graph cleaning, and governance intelligence backend

Extracted entities were canonicalised through conservative normalisation (case/whitespace/punctuation) and assigned deterministic identifiers to support deduplication and stable downstream analysis. A targeted second extraction pass was used to improve coverage based on initial diagnostics. The raw graph was then cleaned through three operations: (i) pruning generic high-degree hub concepts that artificially connect unrelated governance elements, (ii) resolving cross-type duplicates via conservative clustering followed by deterministic LLM adjudication (to correct mis-typed entities and merge true duplicates), and (iii) a final exact-match deduplication pass after type corrections.

The cleaned graph was operationalised as a governance intelligence backend with three integrated components:

1. **Graph store:** a directed multigraph with indices and traversal utilities (entity lookup, neighbour retrieval with predicate/type filters, ego-subgraph extraction, and jurisdiction-scoped queries).
2. **Vector store:** the retained evidence chunks (n = 364) embedded and indexed for semantic retrieval, enabling hybrid evidence access (graph identifies relevant entities/paths; vector search retrieves supporting text).
3. **Query engine:** structured query templates supporting governance pathway tracing (Authority → Instrument → Requirement → Practice → Outcome), jurisdictional comparison, gap analysis (barriers without enabling support), entity explanations, and free-form governance questions grounded in retrieved evidence.

To produce readable outputs while preserving auditability, structured query results were rendered through a deterministic retrieval-augmented generation layer configured to (i) rely only on provided graph structures and evidence excerpts, (ii) cite instruments and sources, (iii) acknowledge uncertainty when evidence is insufficient, and (iv) maintain consistent distinctions among instruments, requirements, practices, and outcomes.

### 3.6 Query templates and governance intelligence layer

The query engine operates through four structured template types, each designed to answer a specific class of governance question. The DfD pathway trace template traverses the graph from a focal practice node (e.g., Design for Disassembly) to identify linked authorities, instruments, requirements, barriers, enablers, and outcome metrics, assessing chain completeness from Authority to Outcome. The jurisdiction coverage template retrieves and compares instrument counts and requirement linkages across nominated jurisdictions (Commonwealth, states, territories). The gap analysis template identifies practice nodes carrying barriers but no supporting instruments or enablers, with severity classified by barrier count. The definitional precision template compares entity and relation counts for generic waste diversion constructs against specific disassembly, salvage, and component reuse constructs to quantify how precisely policy instruments address deconstruction outcomes. All templates are rendered through the retrieval-augmented generation layer, which cites instruments and sources and acknowledges uncertainty when evidence is insufficient.

### 3.7 Extraction yield and quality diagnostics

Pipeline performance was assessed using extraction yield statistics (entities/relations per chunk), graph size before and after cleaning, connectivity diagnostics, and confidence-weight summaries derived from the Keep/Maybe labels. These diagnostics support transparent reporting of what the pipeline extracted from the analysis corpus and enable robustness checks contrasting high-confidence evidence-only outputs with the expanded graph that retains contextual passages.

**Methodology — Code-to-Text Fidelity:** The methodology section is generally accurate but omits several operationally important details that reviewers will expect:

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Code Value** | **Manuscript Says** |
| Chunk window | 1,400 characters | "relatively large chunk window" (unspecified) |
| Chunk overlap | 200 characters | Not stated |
| Similarity threshold | 0.55 (L2, FAISS) | "strict similarity threshold" (unspecified) |
| Within-type clustering cosine | 0.72 | Not stated |
| Cross-type clustering cosine | 0.80 | Not stated |
| LLM models used | Claude Haiku (Pass 1), Claude Sonnet (Pass 2/3, NB4, NB5) | "deterministic LLM" only |
| Raw graph size (pre-cleaning) | 1,748 nodes, 3,929 triples | Not reported |

## 4. Findings

### 4.1. Deconstruction: an overlooked area in Australian circular economy domain

Across the assembled governance corpus (n = 95 documents; 4,313 pages; ~1.42M words), the dominant framing is circular economy and resource recovery, while deconstruction appears as a comparatively weak and unevenly distributed signal. As summarised in Table 1, only 35 of the 95 documents (36.8%) contain any deconstruction-related content based on lexical signal; the GraphRAG analysis corpus comprises 36 documents (including one document that contributes only contextual passages without explicit deconstruction terminology), and the cumulative deconstruction-associated text totals 63,426 words, representing 4.47% of the corpus when weighted by document length. This indicates that deconstruction—despite being operationally central to high-value reuse and component salvage—is not consistently articulated as an explicit governance focus across the broader circular economy evidence base. The temporal distribution also suggests that both the overall corpus and deconstruction-relevant material are concentrated in more recent years.   
  
Table 1. Corpus scope and deconstruction signal summary

|  |  |
| --- | --- |
| Metric | Value |
| Total documents | 95 |
| Total pages | 4313 |
| Total words | 1419553 |
| Docs with deconstruction content | 35 (36.8%) |
| Total deconstruction words | 63426 |
| Deconstruction share of corpus (weighted) | 4.47% |
| Year range (min-max) | 2006-2026 |
| Median year | 2023 |
| Concentration: docs for 50% of deconstruction words | 6 |
| Concentration: docs for 80% of deconstruction words | 13 |
| Max deconstruction intensity (single doc, unweighted) | 37.35% |
| State identified (best-effort) vs Unknown | 21 vs 74 |

Figure 3 shows that most documents appear post-2018, and that deconstruction-positive documents are present but remain a minority within each recent year, reinforcing the interpretation that deconstruction has not been mainstreamed as a core policy object even as circular economy activity increases. Crucially, deconstruction content is not evenly spread across the 35 deconstruction-positive documents; it is highly concentrated in a small subset of sources.



Figure 3. How deconstruction is embedded within the reuse/circular economy evidence base

Figure 3 quantifies this concentration: when documents are ranked by deconstruction-word volume, approximately six documents account for ~50% of all deconstruction-associated words, and thirteen documents account for ~80% (also reported in Table 1). This concentration pattern is consistent with “overlooked core” dynamics: deconstruction receives depth of treatment in a limited number of sources, while a majority of the domain literature advances circular economy objectives using broader (and often less operational) terminology. Finally, Figure 3 explains why deconstruction can be simultaneously “rare in the corpus” yet “dominant in a few documents.” The scatter shows within-document deconstruction intensity (deconstruction words as a percentage of each document’s words), not the corpus share. A small number of documents reach high intensities (maximum 37.35%, Table 1), while many documents sit at or near zero. The dashed reference line at 4.47% (corpus-weighted share) demonstrates that these high-intensity documents are exceptions rather than the norm. Together, Panels B and C in Figure 3 show that deconstruction is present in the evidence base, but specialised and concentrated, supporting the conclusion that it remains under-specified within the broader circular economy governance discourse.

### 4.2. Regulatory Hierarchy and Corpus Composition

Figure 4 illustrates the structural composition of the analysed corpus across four key dimensions. The data reveals a strong reliance on non-government sources (a), with industry, NGOs, and academia contributing the vast majority of deconstruction-relevant content, which contrasts sharply with minimal outputs from federal and state government tiers. Furthermore, the corpus is heavily dominated by reports and guidance toolkits (b), highlighting a lack of formal legislative acts, codes, or standards. Jurisdictional analysis (c) demonstrates a concentration of national and international documents; notably, international guidelines contribute a disproportionately high density of relevant text chunks compared to their document count. Finally, the temporal coverage (d) underscores the recency of this discourse, with over 94% of the literature published since 2020 and distinct peaks occurring in 2022 and 2023.

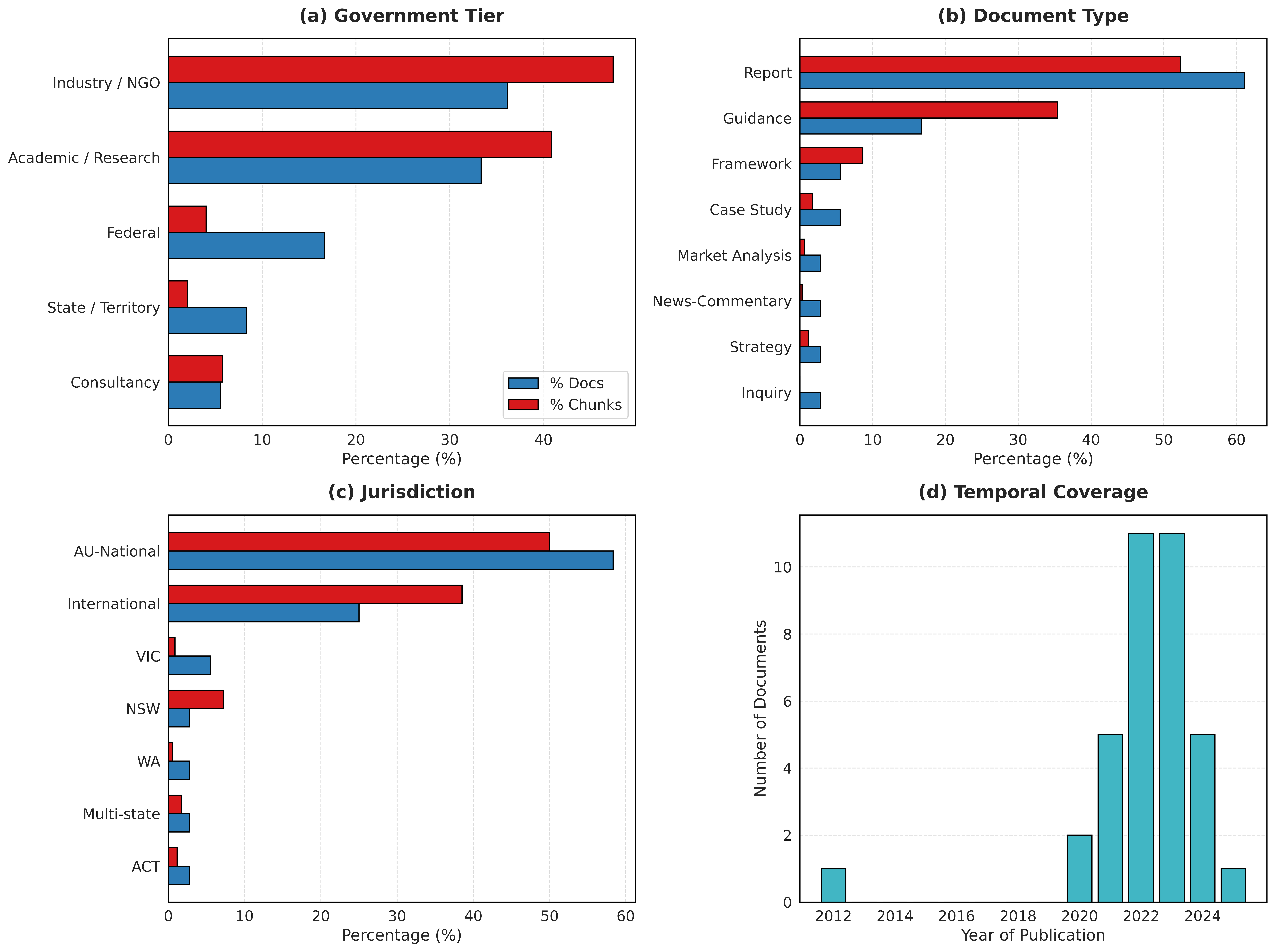


Figure 4. Structural composition of the analysed corpus and the distribution of documents of the retained text chunks

### 4.3. Knowledge Graph Topology: The Structural Emphasis of Governance

To understand how deconstruction is governed, it is first necessary to examine where the regulatory language concentrates. Figure 5 presents the entity and relation type distributions extracted from the analysis corpus, revealing a profound structural imbalance in the Australian governance discourse. Practice nodes dominate the graph (428 nodes; 26.1%), forming the single largest entity type by a substantial margin. This provides the first empirical signal of a "floating discourse": the corpus is heavily preoccupied with describing *what practitioners should do* (e.g., selective dismantling, material tracking) rather than *what governments mandate*.

Conversely, Jurisdiction (41 nodes; 2.5%) and Authority (98 nodes; 6.0%) are structurally underrepresented. The 10:1 ratio of named practices to named jurisdictions demonstrates that governance documents rarely anchor specific circular economy obligations to discrete spatial or political boundaries. Furthermore, Barrier nodes (120 nodes; 7.3%) outnumber both Authority and Enabler nodes (74; 4.5%). This confirms mathematically that there are more documented obstacles to deconstruction in the corpus than there are institutional actors or enabling mechanisms responsible for addressing them.

### 4.4. The Aspirational Register: Facilitation over Mandate

The relations (edges) connecting these entities expose the functional mechanics of the governance landscape. As shown in the right panel of Figure 5, the corpus operates predominantly in an "aspirational register." The predicates INVOLVES (750 edges; 23.3%) and ENABLES (569 edges; 17.7%) collectively account for 41% of all connections. The language of governance relies heavily on associations and facilitations rather than firm mandates. Critically, the ratio of REQUIRES (432 edges) to PROHIBITS (8 edges) is 54:1. This finding is central to explaining the slow uptake of deconstruction in Australia: while instruments occasionally require certain reporting or diversion metrics, they virtually never prohibit practices that are fundamentally incompatible with deconstruction (such as rapid mechanical demolition or the commingling of salvageable materials). The governance framework relies on soft facilitation (ENABLES exceeds REQUIRES by 1.32×) rather than establishing hard regulatory boundaries.

A graph with numbers and a red and blue bar

AI-generated content may be incorrect.

Figure 5. Knowledge graph composition showing the structural emphasis of the governance corpus.

The left panel in Figure 5 details the distribution of entity types, highlighting the dominance of practice and material nodes over jurisdictional anchoring. The right panel illustrates the distribution of semantic relations, demonstrating an "aspirational register" heavily skewed toward enablement and involvement rather than strict requirements or prohibitions (highlighted).

### 4.5. Hub-and-Fragment Structure: The Isolation of Barriers

The network topology further highlights a disconnect between high-level strategies and operational realities, as illustrated in Figure 6. The degree distribution exhibits a strongly right-skewed, near-power-law curve (mean degree 3.93, median 2.0). Only 3.0% of nodes act as highly connected hubs (degree ≥ 20), with "Design for Disassembly" (DfD) emerging as the primary mega-hub (degree 131).

However, this centralization around a few celebrated concepts masks a highly fragmented periphery. Over 41% of nodes are isolates or leaves (degree 0–1). Notably, Barrier nodes possess the lowest mean degree of any entity type (2.02) and the highest isolate rate (21.7%). This indicates that while the industry has converged on a few theoretical solutions like DfD, the actual obstacles practitioners face—such as conflicting local permits or insurance premiums on reused steel—are experienced in silos. They are documented only once, in single sources, lacking the convergent, systemic recognition required for targeted policy intervention.

### 4.6. Instrument Subtypes and the Statutory Vacuum

While Instrument is the second most common entity type (245 nodes; 14.9%), a sub-classification of these nodes reveals a critical statutory vacuum (Figure 6, right panel). Only 13.1% of the named instruments are classifiable as statutory acts, regulations, or building codes. The remaining 86.9% comprise voluntary strategies, best-practice guidelines, and roadmaps. Further semantic tracing reveals that many of the high-degree statutory instruments cited within the Australian corpus are actually imported (e.g., the EU Waste Framework Directive, UK Construction Products Regulation). When filtered strictly for domestic statutory instruments, the list is reduced to generic environmental protection acts and the National Construction Code, which heavily favour generic waste diversion over explicit deconstruction mandates. This aligns with the "aspirational register" finding, confirming that the instruments doing the heavy lifting in the Australian context are guidance documents lacking enforceable legal mechanisms.

As in Figure 6, the left panel displays the log-scaled node degree distribution, revealing a hub-and-fragment structure where the vast majority of nodes (including key barriers) remain isolated or loosely connected. The right panel classifies the identified governance instruments, exposing a profound statutory vacuum where formal legislative acts and codes make up only a small fraction of the guiding literature.

A graph of a diagram

AI-generated content may be incorrect.

Figure 6. Network topology and regulatory mechanisms.

## 4.6. Platform Demonstration: Query-Driven Policy Intelligence

To transition the GraphRAG pipeline from a static analytical script into an accessible governance intelligence tool, a standalone interactive presentation layer was developed. The computational backend—comprising the cleaned NetworkX property graph and the semantic text chunks—was preserved via serialization. A lightweight frontend web application was then deployed using the Gradio framework directly over this serialized state. Upon initialization, the application dynamically rehydrates the knowledge graph, reconstructs the high-dimensional vector store in-memory (using the all-MiniLM-L6-v2 embedding model), and connects to an external Large Language Model (claude-haiku-4-5) constrained by a strict, zero-hallucination system prompt. This architecture allows end-users to submit natural-language policy queries. The system executes a multi-step retrieval process: (i) semantic vector search to identify relevant regulatory excerpts, (ii) graph traversal to extract the relational neighbourhood (e.g., connected Authorities, Instruments, and Barriers) of any entities mentioned in the query, and (iii) deterministic synthesis of the final answer, explicitly citing the retrieved nodes and documents. To demonstrate the platform's capacity for query-driven policy intelligence, an operational case study was executed against the Australian circular construction corpus. Figure 7 presents a structured summary of four representative queries synthesized by the GraphRAG system, demonstrating its ability to perform pathway tracing, multi-hop jurisdictional comparisons, and gap analyses.

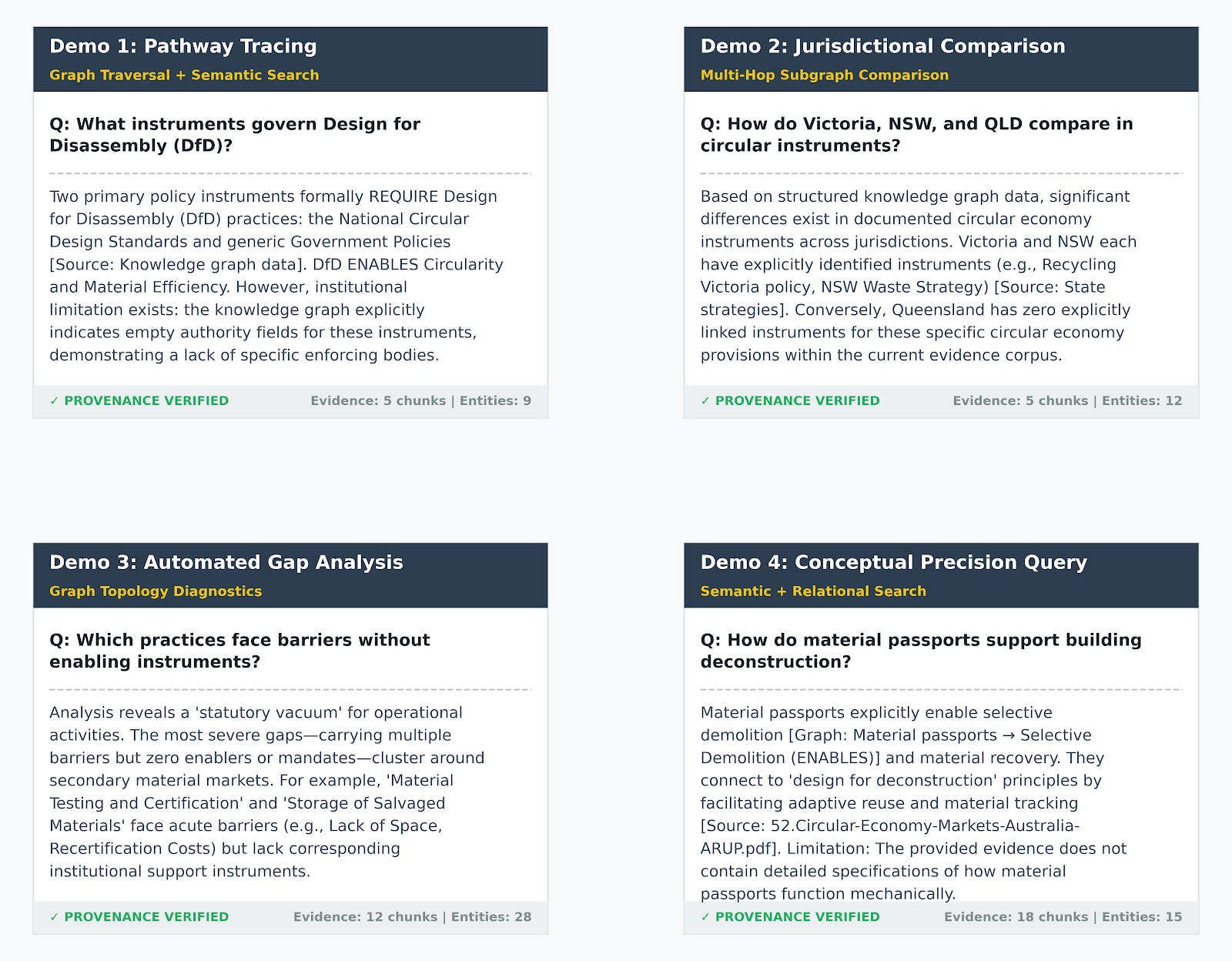


Figure 7. Demo cards

To test the platform's semantic precision and resistance to generative hallucination, a specific query asked, *"How do material passports support building deconstruction?"* The system successfully isolated the operational mechanics of this intervention using both structural edges and semantic text chunks. As shown in Figure 7, the interactive interface generated an answer that accurately reported material passports directly *ENABLE* "Selective Demolition" and "Material Recovery" practices, linking these to specific downstream assets. Furthermore, the platform successfully contextualized the operational scope, citing a specific retrieved document (e.g., *52.Circular-Economy-Markets-Australia-ARUP.pdf*) to connect material passports with adaptive reuse. Most importantly, the generated response included a strictly generated limitation clause: *"The provided evidence does not contain detailed specifications of how material passports function mechanically in deconstruction processes."* Because the LLM was bounded exclusively by the semantic chunks and graph triples provided by the retrieval pipeline, it correctly refused to extrapolate or hallucinate technical mechanisms not present in the underlying PDF corpus. A core feature of this interactive deployment is the preservation of analytical transparency. For every generated response, the frontend exposes an audit trail (visible in the expanded accordion in Figure 7) containing the raw JSON payload of the retrieved graph context and the exact semantic evidence passages fed to the LLM. This ensures that all policy intelligence generated by the platform remains fully auditable, source-grounded, and verifiable by human researchers, overcoming the "black box" limitations of standard generative AI applications in regulatory analysis.

## Discussion

## 5. Conclusions

Several limitations bear on the scope and interpretation of findings. First, the corpus is bounded by Australian instruments available and ingested at the time of retrieval; instruments not captured through the search workflows are not represented in the knowledge graph. Second, “Maybe” passages are treated as contextual evidence with a lower confidence weight (0.6), and findings that rely substantially on these passages should be interpreted with appropriate caution. Third, provenance completeness varies across the corpus: page numbers are recorded where PDF extraction allows, but some scanned or structurally complex documents yield chunk-level provenance only. Future research should extend the corpus systematically across all Australian jurisdictions, introduce active human validation of a stratified sample of extracted relations, and explore temporal updating of the knowledge graph as new policy instruments are issued.

## References