# 1. Research Aim and Objectives

## 1.1. Research Aim

This study aims to develop and validate a computational governance intelligence platform for systematically mapping, querying, and critically evaluating the Australian regulatory ecosystem governing building deconstruction and material salvage. Specifically, it combines large language model (LLM)-driven information extraction with knowledge graph (KG) construction and retrieval-augmented generation (RAG) to transform an unstructured corpus of federal, state, and local policy documents into a structured, auditable, and queryable **Governance Topography**. The platform enables automated identification of legislative disconnects, jurisdictional fragmentation, and governance gaps that hinder the transition from conventional demolition to a circular built environment.

## 1.2. Research Objectives

***Objective 1: Constructing a Multi-Stage Computational Pipeline for Regulatory Corpus Processing.*** To design, implement, and validate a reproducible six-stage computational pipeline that transforms heterogeneous government PDF documents (n=36) into a structured knowledge graph. This pipeline integrates multi-engine text extraction (layout-aware parsing, robust encoding resolution, and optical character recognition), semantic chunking with domain-specific keyword gating, high-dimensional vector retrieval, and LLM-based precision filtering to isolate regulatory clauses specific to structural disassembly, selective demolition, and material salvage from broader waste management narratives.

***Objective 2: Developing an Objective-Aligned Knowledge Graph Schema for Deconstruction Governance.*** To define and populate a purpose-built knowledge graph schema comprising ten entity classes (Instrument, Authority, Jurisdiction, Requirement, Practice, MaterialAsset, Stakeholder, Barrier, Enabler, OutcomeMetric) and eleven governance-mechanism relation types that reconstruct “who governs what, where, and how.” This schema supports direct traceability from every extracted statement back to its source document, page number, and verbatim evidence excerpt, enabling end-to-end auditability.

***Objective 3: Analysing Semantic and Definitional Gaps in Deconstruction Policy.*** To deploy the governance intelligence platform to critically examine the extent to which “deconstruction” and “salvage” are formally recognised as distinct regulatory activities within Australian policy. This involves quantifying the prevalence of generic “waste diversion” language (which encourages crushing and downcycling) versus specific mandates for “disassembly” and “component reuse,” and mapping the definitional boundaries across jurisdictions.

***Objective 4: Mapping the Institutional Governance Network and Jurisdictional Fragmentation.*** To use structured graph traversal queries to map the ecosystem of institutional actors—including Environmental Protection Authorities (EPAs), sustainability bodies, and local planning authorities—and to characterise the structural relationships (or absence of coordination) between agencies responsible for waste minimisation and those responsible for building standards. This includes automated jurisdictional comparison across Australia’s three tiers of government.

***Objective 5: Synthesising Governance Gaps and Regulatory Barriers Through Automated Gap Analysis.*** To systematically identify practice areas where documented barriers (conflicting permit requirements, risk-averse procurement policies, absence of certification pathways for reused materials) outweigh or entirely lack corresponding enabling instruments or policy support. The gap analysis classifies governance deficits by severity—distinguishing practices with zero institutional support from those with partial but insufficient coverage.

## 1.3. Expected Contribution

This research makes three distinct contributions. **First**, it provides the first computational regulatory analysis of **deconstruction** as a specific subset of the circular economy in Australia, moving beyond broad “waste” narratives to offer a diagnostic assessment of specific legislative gaps. **Second**, it demonstrates a replicable methodology for transforming unstructured policy corpora into structured, queryable knowledge graphs using LLM-driven extraction with full evidence provenance—an approach transferable to other regulatory domains and jurisdictions. **Third**, it delivers an operational governance intelligence backend that combines graph traversal with retrieval-augmented generation, enabling policymakers and researchers to pose structured queries against the regulatory landscape and receive grounded, source-cited answers.

# 3. Research Methods and Design

## 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.

A diagram of a diagram

AI-generated content may be incorrect.

Figure . Pipeline overview

### 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. However, the **analysis corpus** for knowledge-graph construction was defined after precision filtering: **36 of the 95 documents** contained passages directly relevant to structural disassembly, salvage, and/or regulatory barriers/enablers for reuse, and therefore contributed evidence to the graph. The remaining **59 documents** were excluded from graph construction because they primarily addressed generic waste management or demolition narratives without substantive deconstruction-specific content. The final analysis corpus comprised **364 high-relevance text chunks** drawn from these 36 documents.

### 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. 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, PRODUCES). 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, page number, chunk identifier/hash). Relations were retained only when both endpoints were explicitly extracted from the same chunk, reducing speculative linkage.



Figure . Knowledge-graph schema

### 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 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.

# 4. Findings

*This section presents the results of deploying the governance intelligence platform against the Australian deconstruction policy corpus. Findings are organised around the four analytical dimensions enabled by the platform’s query engine.*

## 4.1. Regulatory Hierarchy and Corpus Composition

*➤ Report the composition of the 36-document corpus: breakdown by government tier (federal, state, local), document type (Act, Bill, Code, Standard, Strategy, Report), and temporal coverage. Present the entity type distribution from the knowledge graph.*

**[Figure 3]** *Summary statistics dashboard: (a) entity type distribution, (b) relation type distribution, (c) node degree distribution (log-scale), (d) platform metrics table.*

*➤ Discuss how the entity distribution reveals the structural emphasis of Australian governance: what entity types dominate, what is underrepresented, and what this implies about regulatory priorities.*

## 4.2. Governance Pathway Analysis: Design for Disassembly

*➤ Present the results of the governance pathway trace for Design for Disassembly (DfD) as the primary case study. Report the number of instruments, requirements, enablers, barriers, and outcomes connected to DfD in the knowledge graph. Discuss the governance chain: which authorities issued which instruments, and which jurisdictions they apply in.*

**[Figure 4]** *DfD governance pathway ego network showing instruments, requirements, barriers, and outcomes connected to Design for Disassembly practice node.*

*➤ Analyse the pathway for completeness: does the chain from Authority → Instrument → Requirement → Practice → Outcome hold, or are there broken links? Identify which jurisdictions have stronger governance chains and which have gaps.*

## 4.3. Jurisdictional Comparison

*➤ Present the jurisdictional comparison results for Victoria, New South Wales, and Queensland (and other states as available). Report the number of instruments per jurisdiction, shared instruments across all three, and unique instruments per jurisdiction. Discuss the coverage matrix.*

**[Figure 5]** *Jurisdictional coverage heatmap: (a) stacked bar chart of entity connections by jurisdiction, (b) binary heatmap of instrument applicability across jurisdictions.*

*➤ Discuss the implications of the Queensland attribution gap identified in the knowledge graph cleaning stage. Analyse whether jurisdictional fragmentation is primarily in instrument coverage, enforcement mechanisms, or both. Compare federal aspirational language versus state-level statutory reality.*

## 4.4. Gap Analysis: Barriers Without Enabling Instruments

*➤ Present the gap analysis results: total practices analysed, number of gaps identified, and number of high-severity gaps (practices with barriers but zero instruments or enablers). List the most critical gaps and their associated barriers.*

**[Figure 6]** *Gap analysis matrix: barriers versus instruments and enablers per practice, with high-severity gaps (zero support) highlighted.*

*➤ Interpret the gap analysis in the context of regulatory inertia: which barriers are legislative (conflicting permits, missing certification), which are institutional (lack of coordination between agencies), and which are market-based (risk-averse procurement)? Discuss how the severity classification (high vs. medium) maps to policy actionability.*

## 4.5. Semantic and Definitional Gaps

*➤ Present findings on how ‘deconstruction’ and ‘salvage’ are defined (or not) in the policy corpus. Report the prevalence of generic ‘waste diversion’ language versus specific ‘disassembly’ and ‘component reuse’ mandates. Use entity and relation counts from the knowledge graph to quantify definitional precision.*

*➤ Discuss the implications: if policy instruments refer only to ‘waste diversion’ without distinguishing between crushing/downcycling and selective disassembly/reuse, what are the practical consequences for deconstruction practitioners?*

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

*➤ Present the four demo queries and their Claude-synthesised answers as evidence that the platform functions as an operational governance intelligence tool. Summarise each demo: (1) DfD pathway, (2) jurisdictional comparison, (3) gap analysis, (4) material passports and deconstruction. Emphasise that every claim in the synthesised answers traces back to source documents.*

**[Figure 7]** *Demo Q&A cards: 2×2 grid showing four formatted query–answer pairs with source citations and structured data badges.*