

KnowledgeSlot Roadmap

Purpose: Curate, structure, and prepare domain knowledge content for ingestion into the Knowledge Slot of a Cosolvent marketplace deployment.

Date: March 6, 2026 **Author:** Mustafa Uzumeri

1. What is the Knowledge Slot?

The Knowledge Slot is a **sponsor-curated reference library** that provides domain knowledge to a Cosolvent marketplace deployment. It is one of five architectural "slots" in the Cosolvent framework:

Slot	What it holds	Who curates it
Context Slot	Participant-supplied documents (uploads, profiles)	Participants
Intelligence Slot	AI model configuration, provider routing, prompts	Admin / Developer
Knowledge Slot	Sponsor-curated domain reference material	Marketplace sponsor
Agent Slot	Brokerage agent configuration (personas, rules)	Admin
MCP Slot	External data source / tool connectivity	Admin

Terminology Evolution

The Knowledge Slot concept has evolved through several design iterations — from a general-purpose document store in early prototypes, to an architecturally distinct repository for sponsor-curated, domain-authoritative reference material. The key distinction: **participant documents** (Context Slot) follow a three-layer privacy model and are self-service. **Reference documents** (Knowledge Slot) are sponsor-curated, progressively built, and authoritative. The two never mix in retrieval.

2. Design Principles

2.1 — Separation from Participant Documents

Reference documents are stored separately from participant-uploaded files. This ensures:

- Reference material is not contaminated by participant uploads
- Participant privacy controls do not apply to reference material (it is sponsor-curated, not user-submitted)
- Retrieval pipelines can scope queries to one or both stores

2.2 — Vertical-Specific Metadata Schema

The tag vocabulary for reference documents is defined by the vertical deployment, not hardcoded. Each marketplace defines its own metadata schema appropriate to the domain.

Examples by vertical:

Vertical	Example Metadata Tags
Agricultural trade	Origin region, destination country, product category, document type, trade corridor, issuing body
Remote mental health	Jurisdiction, insurance provider, service type, clinical area, license type, regulatory body
Specialty manufacturing	Material class, standards body (ISO / ASTM / DIN), export control regime, process type

2.3 — Document Curation Workflow

Administrators can upload, tag (using the vertical's metadata schema), describe, and version reference documents. The sponsor progressively builds this library over time. Documents are chunked, embedded, and immediately available for retrieval.

2.4 — Metadata-Filtered Vector Search

Retrieval uses metadata pre-filters *before* vector similarity ranking. This avoids physical partitioning while achieving context-appropriate scoping. Documents can carry multiple tags (e.g., a guide to international trade terms is tagged with every destination country it applies to), and cross-corridor queries work naturally.

2.5 — Automatic User-Context Scoping

The system injects the participant's metadata (country, role, product interests, active deal corridors) as implicit retrieval filters. Users don't need to specify "show me Philippines regulations" — they just ask their question, and the system scopes retrieval based on what it already knows about them.

2.6 — Domain Q&A Integration

The chatbot supports a "domain knowledge" mode where it answers from the reference library, distinct from participant-to-participant messaging or deal-scoped conversation. Answers are sourced — each response cites which reference documents informed the answer.

2.7 — Vertical-Supplied Prompts

The Knowledge Slot's chat behaviour is configured per-vertical: the deployment defines the AI persona, citation style, and scope boundaries (e.g., "only answer from reference material; say 'I don't have this information' otherwise").

3. Cross-Slot Architectural Guardrails

Three design constraints preserve the path to cross-slot intelligence:

1. **Same embedding model across all vector stores.** The reference library and participant embeddings must use the same embedding model. If they diverge, cross-collection similarity search becomes impossible without re-embedding.

2. **Shared metadata vocabulary.** Concepts that appear in both participant profiles and reference documents — geography, product categories, certification types — must use the same controlled vocabulary.
3. **Composable retrieval interface.** The retrieval layer returns results in a uniform format regardless of which store they came from, and accepts a source parameter (participants, reference library, or both) so that future callers can request merged, cross-collection results.

4. What This Project Produces

This project curates the **content** side of the Knowledge Slot — the reference documents, domain schemas, and metadata that a marketplace sponsor would load into the reference library.

4.1 — Deliverables

Deliverable	Description	Status
Converted reference documents	Source documents (PDF, HTML, CSV, XLSX) converted to clean Markdown	✓ Complete
Domain schema	Structured vocabulary extracted from reference documents (YAML)	✓ Complete
Schema analysis results	LLM-generated proposals for schema additions and refinements	✓ Complete
Analysis prompt template	Editable LLM prompt for schema extraction	✓ Complete
Provenance metadata	Source URL, acquisition method, and timestamps for every document	✓ Complete
Metadata extraction prompt	Editable LLM prompt for document metadata extraction	✓ Complete
Metadata tag vocabulary	Vertical-specific tags for the reference metadata schema	✓ Complete
Participant role map	Industry roles mapped to Cosolvent supply/demand/facilitator categories	✓ Complete
Referenced standards inventory	Standards incorporated by reference — candidates for future ingestion	✓ Complete
Process recipe	Documented workflow for curating additional content	✓ Complete

4.2 — Current Content

The initial content vertical is **agricultural commodity trade** (grain trading). The first reference document — a standard international grain trading contract — has been fully converted, schema-extracted, and analyzed. This document is representative of the types of material that would be ingested: industry-

standard contracts, trade rules, grading standards, and regulatory frameworks published by trade associations and government agencies.

Additional documents have been identified for future ingestion, including supplementary trade rules (weighing, sampling, arbitration, analysis methods, fumigation, insurance), as well as government grading standards from national agencies.

5. Roadmap Phases

Phase 1 — Foundation Complete

Goal: Establish tooling and process; produce first domain schema from a reference contract.

- PDF-to-Markdown conversion pipeline
- URL-to-Markdown conversion pipeline
- First reference document converted and schema extracted
- Participant roles mapped to Cosolvent categories
- Curation process documented
- Web GUI for document ingestion and conversion
- LLM-assisted schema analysis via OpenRouter
- Editable prompt template for schema extraction
- GUI for reviewing LLM-generated schema proposals
- Provenance tracking — source URL, acquisition method, timestamps for every document
- Smart URL detection — auto-detects HTML vs. downloadable files and routes correctly
- YAML frontmatter injection — every output includes source URL for downstream chunking
- LLM-assisted metadata extraction — imputes org, author, date, doc type for locally-uploaded files
- Editable metadata extraction prompt
- CSV and Excel (.xlsx) tabular data conversion — renders spreadsheets as Markdown tables with full provenance
- Docker packaging for deployment

Phase 2 — Schema Enrichment

Goal: Process additional contracts and standards to broaden and refine the domain schema.

- Acquire additional industry contracts covering different trade configurations (FOB terms, specific commodities)
- Convert and run schema analysis on additional contracts
- Review LLM-generated proposals and merge approved additions into the domain schema
- Merge schemas — identify common entities, resolve conflicts, note configuration-dependent variations
- Process referenced industry standards (weighing, sampling, arbitration, analysis, fumigation, insurance)
- Add government grading standards
- Define metadata tag vocabulary for the grain trading vertical

Phase 3 — Ingestion Preparation

Goal: Prepare content in the format required by the Cosolvent reference library.

- Define chunking strategy that preserves clause-level coherence
- Tag each document chunk with vertical-specific metadata
- Leverage provenance metadata (source URL, source type, document title) as chunk-level citation data
- Generate embeddings for reference document chunks
- Create seed data scripts for the reference library
- Write domain Q&A system prompts for the grain trading vertical
- Define the reference metadata schema configuration for grain trading

Phase 4 — Integration Testing

Goal: Validate that curated content works correctly when loaded into a Cosolvent instance.

- Load curated content into a Cosolvent test instance
- Verify metadata-filtered vector search returns relevant results
- Test domain Q&A with grain trading questions
- Verify user-context scoping (e.g., a buyer in Japan gets Japan-relevant regulations)
- Test cross-collection retrieval (reference library + participant data)
- Validate facilitator role recommendations based on schema

Phase 5 — Additional Verticals

Goal: Demonstrate that the curation process generalises beyond grain trading.

- Identify a second vertical (e.g., remote mental health, specialty manufacturing)
- Apply the curation recipe to the new vertical's reference documents
- Extract a domain schema for the new vertical
- Validate that the same tooling and process works across verticals

6. Dependencies

Dependency	Status	Impact
Reference library table schema in Cosolvent	<input type="checkbox"/> Not yet implemented	Phase 3 ingestion format depends on this
Reference metadata schema configuration	<input type="checkbox"/> Not yet implemented	Phase 3 metadata tagging depends on this
Domain Q&A chat mode in Cosolvent	<input type="checkbox"/> Not yet implemented	Phase 4 testing depends on this
Embedding model choice	<input checked="" type="checkbox"/> Established	Phase 3 embeddings must match this model
Composable retrieval interface	<input type="checkbox"/> Not yet implemented	Phase 4 cross-collection testing depends on this

7. Relationship to Thin Market Theory

The Knowledge Slot operationalises several concepts from the DeeperPoint thin market framework:

Concept	Challenge	How KnowledgeSlot Addresses It
Authoritative Information Availability	Fragmented, hard-to-find reference material thins markets	Centralises authoritative documents in a searchable, AI-accessible library
Curating & Distributing Authoritative Information	Reference library creation, contextual delivery, cross-reference mapping	Curation workflow + metadata-filtered retrieval + user-context scoping
AI-Curated Authoritative Information	RAG over reference library, cross-reference mapping, natural language Q&A	Domain Q&A mode, vertical-supplied prompts, citation-backed answers
Reducing Information Asymmetry	Knowledge imbalance between buyer and seller	Shared access to authoritative standards levels the playing field
Simplifying Regulatory Compliance	Compliance databases, automated checking	Reference library tagged by jurisdiction enables compliance-aware retrieval

8. Open Questions

- 1. Schema versioning:** How should the domain schema evolve as new documents are processed? Additive-only, or refinement allowed?
- 2. Cross-document conflicts:** What happens when two industry contracts define the same entity differently (e.g., different tolerance rules for FOB vs. CIF terms)?
- 3. Chunking strategy:** What chunking approach preserves clause-level coherence while staying within embedding model context limits?
- 4. Multi-vertical schema inheritance:** Should schema formats support inheritance (a base trading schema extended by corridor-specific schemas)?
- 5. Sponsor curation UX:** What does the admin workflow look like for a non-technical sponsor uploading and tagging documents?
- 6. Update monitoring:** How should the system handle updates to reference documents (e.g., a trade association publishes a new edition of a standard contract)?