MarketForge Roadmap

> **Purpose:** The convergence of three tools — KnowledgeSlot, ClientSynth, and Cosolvent — into a single workflow for building, demonstrating, and launching AI-assisted marketplace platforms for thin markets.
>
> **Date:** March 6, 2026 **Author:** Mustafa Uzumeri

1. What is MarketForge?

MarketForge is the integration of three existing tools into a single workflow that produces a functioning digital twin of a proposed thin market. Given a market description and domain reference materials, MarketForge:

1. **Curates domain knowledge** — ingests trade regulations, quality standards, contract templates, and procedural guides (KnowledgeSlot)
2. **Generates marketplace configuration** — extracts participant types, profile schemas, matching prompts, and deal parameters from the domain knowledge
3. **Synthesizes a realistic population** — generates demographically plausible, economically coherent, culturally appropriate synthetic participants (ClientSynth)
4. **Stands up a configured marketplace** — compiles and populates a Cosolvent instance with the reference library, configuration, and synthetic population
5. **Runs simulation scenarios** — executes matching, Deal Brief assembly, facilitator search, and Handoff Artifact generation against the synthetic population
6. **Validates the business design** — the sponsor uses the digital twin as a sandbox to pretest value-added services with real service providers against the synthetic population
7. **Transitions to a real market** — synthetic participants are retired entirely and real participants are recruited by the sponsor into a production instance

The output progresses from a **clickable, demonstrable prototype** (Phases 1–6) through a **business design sandbox** (Phase 7) to a **live marketplace** (Phase 8).

2. The Three Tools and Their Roles

Tool	What it produces	Role in the digital twin	Current state
KnowledgeSlot	Reference documents, domain metadata schemas, provenance records, extraction prompts	**The rules of the game** — regulations, standards, and contracts that govern the market	Working: document conversion, provenance tracking, metadata extraction. Planned: ingestion-ready output format for Cosolvent
ClientSynth	Synthetic participant populations with	**The players** — who would	Working: schema designer, AI generation, export

Tool	What it produces	Role in the digital twin	Current state
	configurable schemas, demographic distributions, and behavioral parameters	participate, with what capabilities and needs	system. Planned: Cosolvent export format, domain-aware vocabulary
Cosolvent	Marketplace infrastructure — configuration, semantic matching, communication, facilitator types, admin oversight, prompt management	**The arena** — the operational mechanics that connect participants and facilitate deals	Working: YAML configuration, dynamic profiles, vector search, admin tools, prompt management. Planned: Deal entity, Handoff Artifact, Knowledge Slot integration

What makes this convergence non-obvious

The three tools were developed independently with different goals. The insight is that their outputs form a **closed triangle** — each tool's output is another tool's input:

- **KnowledgeSlot** produces a domain schema and reference library
- The domain schema informs **ClientSynth's** generation vocabulary, making synthetic participants use real-world terminology rather than generic placeholders
- The domain schema also drives **Cosolvent's** marketplace configuration — participant types, profile fields, matching prompts, and deal parameters
- **ClientSynth** populates the **Cosolvent** instance with a synthetic participant population

The missing connector — domain schema informing ClientSynth's generation vocabulary — is the piece that closes the loop and makes the system self-consistent.

---

3. Integration Points

Already designed (not yet built)

Connection	What it does
KnowledgeSlot → Cosolvent	Reference documents loaded into the marketplace reference library
ClientSynth → Cosolvent	Synthetic participants exported to Cosolvent participant records
Cosolvent config → ClientSynth	Marketplace configuration drives population schema and counts

New connections needed (not on any existing roadmap)

Connection	What it enables	Priority
KnowledgeSlot → ClientSynth	Domain schema constrains generation vocabulary — synthetic participants use real-world terminology	**Critical**

Connection	What it enables	Priority
KnowledgeSlot → Cosolvent config	Domain schema drives automatic generation of participant types, profile fields, and matching prompts	**Critical**
Cosolvent simulation → Analytics	Deal outcomes, match quality, and market dynamics feed into reporting for sponsors and investors	**Medium**

4. The End-to-End Workflow

Phase 1: Market Description

Input: A natural-language market description, plus reference materials.

Activities:

1. Sponsor describes the market in natural language
2. Sponsor uploads or points to reference materials (regulatory documents, trade agreements, quality standards, contract templates)
3. AI assists in refining the market description through structured dialogue — eliciting participant types, geographic scope, regulatory jurisdictions, product categories, and deal structure

Output: Structured market brief and uploaded reference documents

---

Phase 2: Knowledge Curation

Tool: KnowledgeSlot

Activities:

1. Reference documents are converted from PDF, HTML, spreadsheets, and other formats to clean, structured text
2. Provenance metadata is extracted or inferred (organization, date, document type, jurisdiction)
3. Domain schema is extracted — a vocabulary defining participant type taxonomy, profile field definitions, product and service attributes, geographic scope, regulatory context, quality grading systems, and deal parameters
4. Documents are chunked, embedded, and tagged with the domain schema's metadata vocabulary
5. Human reviews and refines the domain schema

Output:

- Domain schema — the vertical's vocabulary and structure
- Reference library — chunked, embedded, metadata-tagged reference documents
- Extraction prompts tuned for this domain

Phase 3: Marketplace Configuration

Tool: Cosolvent configuration generator (new component)

Activities:

1. Domain schema is transformed into a marketplace configuration file
2. Matching prompts are generated from the domain schema (what dimensions matter for semantic matching in this vertical)
3. Deal Brief template is generated (vertical-specific sections and field mappings)
4. Facilitator role slot definitions are generated (what facilitator types exist, which are required or optional per deal type)
5. Human reviews and adjusts the generated configuration

Output:

- Complete Cosolvent deployment configuration
- Matching, extraction, brokerage, and rationale prompts
- Handoff Artifact structure definition

Phase 4: Population Synthesis

Tool: ClientSynth

Activities:

1. ClientSynth reads the marketplace configuration and domain schema
2. Population specification is derived: participant types and counts, profile field distributions per type, vocabulary constraints from domain schema, economic coherence rules, and geographic distribution
3. Synthetic participants are generated with full profiles
4. Human reviews sample profiles for plausibility

Output:

- Synthetic participant records conforming to the marketplace schema
- Generation parameters, distributions used, and vocabulary sources

---

Phase 5: Assembly and Deployment

Tool: Cosolvent deployment pipeline

Activities:

1. Reference library is loaded with metadata tags
2. Marketplace configuration is compiled
3. Synthetic participants are loaded with properly structured profiles
4. Participant embeddings are generated and indexed for semantic search
5. System prompts are loaded from the generated prompt set
6. Health checks confirm the instance is operational

Output: Running Cosolvent instance with populated marketplace

---

Phase 6: Simulation and Demonstration

Activities:

1. **Gallery browsing** — browse synthetic participant profiles as different participant types
2. **Match generation** — run semantic matching to find buyer-seller pairs
3. **Deal Brief assembly** — initiate deals between matched pairs and collaboratively assemble the Deal Brief
4. **Facilitator search** — trigger deal-triggered facilitator discovery for required role slots
5. **Handoff Artifact generation** — assemble the complete deal record from the collaborative process
6. **Knowledge Slot Q&A** — query the reference library from a participant's perspective
7. **Market dynamics reporting** — aggregate match quality, deal completion rates, facilitator utilization, and other metrics
8. **Demo Mode preparation** — configure for public showcase

Output:

- Demonstrable prototype with realistic market activity
- Market dynamics report with quantitative evidence
- Public demo instance (read-only, zero cost per visit)

Phase 6a: Demo Mode — Public Showcase

Once Phase 6 has produced a working digital twin, the instance can be switched to **Demo Mode** — a read-only, publicly accessible showcase that lets visitors experience the marketplace as a synthetic participant.

Key design:

- All database writes are blocked — data is read-only
- Live AI prompts are replaced with pre-computed results, eliminating per-visitor cost and security exposure
- Visitors are assigned a synthetic persona (not a real account) — no data is collected or persisted
- Sessions evaporate when the visitor leaves

Visitor experience:

1. Visitor selects a role (e.g., grain exporter, mill/processor, shipping provider, inspector, trade finance)
2. System randomly assigns a synthetic participant of that type as the visitor's persona
3. Visitor navigates the marketplace as their assigned persona with curated "Ask about this" buttons at each screen
4. Pre-computed responses to curated questions are returned instantly, at no per-visit cost

Pre-computation: All match results, match rationale, sample Deal Briefs, Knowledge Slot Q&A, facilitator recommendations, market analytics, and peer comparisons are generated once (a one-time AI cost) and cached. Every subsequent visitor interaction is a cache lookup.

Phase 7: Value-Added Service Pretesting

This phase uses the digital twin as a **business design sandbox**. The marketplace holds synthetic participants, but the services being tested may involve real-world actors — a real bank, a real logistics provider, a real compliance advisor.

Activities:

1. **Sample fulfillment testing** — Using deal parameters from synthetic deals, test whether a physical sample can move through the logistics chain. This validates that the Handoff Artifact contains enough information to drive real-world logistics.
2. **Market analytics product testing** — The digital twin generates match data, deal patterns, facilitator utilization, corridor traffic, and pricing signals. Test whether these analytics have standalone commercial value.
3. **Trade finance integration** — Show a completed Deal Brief to a bank's trade finance desk. Test whether the document is sufficient to initiate a letter of credit assessment.
4. **Compliance and regulatory advisory** — Test the Knowledge Slot Q&A system as a standalone advisory service.
5. **Insurance integration** — Present synthetic deal parameters to a cargo insurer. Test whether they can price a policy from the Deal Brief's data.
6. **Platform UX testing with real service providers** — Have real facilitators use the platform with synthetic principals.

What makes this distinctive: Synthetic participants provide the context — the deals, the corridors, the commodity flows. But the service providers being tested are real. Synthetic users are clearly labelled as synthetic throughout.

Output:

- Service integration test results
- Refined Deal Brief templates incorporating real-world feedback
- Market analytics samples demonstrating subscription product viability
- Business model validation: which value-added services generate willingness-to-pay

---

Phase 8: Market Launch — Synthetic to Real Transition

This is the point at which the digital twin has served its purpose and the sponsor transitions to a **real marketplace with real participants**.

> ⚠️ **Critical ethical constraint:** Synthetic users must **never** coexist with real users in the same marketplace instance. The transition is a complete cutover, not a gradual blend. Mixed use would deceive real participants about who they are interacting with and would undermine the trust the platform exists to build.

The clean cutover:

1. **Preserve configuration** — Marketplace configuration, prompts, Deal Brief templates, facilitator role definitions, and reference library carry forward unchanged
2. **Archive synthetic data** — All synthetic participant records, synthetic deals, and synthetic Deal Briefs are archived and then removed from the production instance. The production database starts with zero participants.

3. **Preserve the reference library** — Knowledge Slot content (reference documents, metadata, embeddings) carries forward as sponsor-curated domain knowledge
4. **Preserve service integrations** — Any integrations tested in Phase 7 carry forward
5. **Deploy production infrastructure** — Production hosting, monitoring, backup, and security
6. **Begin real participant recruitment** — The sponsor's operational responsibility

The sponsor's ongoing responsibilities after cutover:

Responsibility	Why it's the sponsor's
Participant recruitment	Requires domain-specific go-to-market strategy and industry relationships
Customer support	Requires domain expertise and human judgment
Content maintenance	Keeping the reference library current as regulations change
Platform operations	Hosting, monitoring, backups, security, uptime
Business development	Building relationships with service providers
Regulatory compliance	Jurisdiction-specific legal expertise

DeeperPoint / MarketForge continues to provide:

- Framework updates and bug fixes
- Architectural guidance on scaling and new module integration
- New vertical digital twins for adjacent corridors or markets
- Knowledge curation support for updating the reference library

---

5. What Cannot Be Automated

Even in the fully operational Forge, certain tasks require human judgment:

Task	Who	When
Market selection and scoping	Sponsor / domain expert	Phase 1
Reference document curation	Domain expert	Phase 2
Domain schema review	Domain expert	Phase 2
Marketplace config review	Domain expert	Phase 3
Population plausibility check	Domain expert	Phase 4
Simulation interpretation	Analyst / sponsor	Phase 6
Demo presentation	Sponsor / team	Phase 6
Service integration negotiation	Sponsor	Phase 7
Participant recruitment	Sponsor	Phase 8
Cutover decision	Sponsor	Phase 7→8

6. Effort and Time Estimates

The effort required varies with market vertical complexity. These estimates assume the Forge toolchain is fully operational.

Market Complexity Tiers

Tier	Example verticals	Profile
Tier 1 — Simple	Local farmers' market, peer-to-peer equipment rental, single-commodity domestic trade	2 participant types, 1 jurisdiction, commodity product, simple deals
Tier 2 — Moderate	Cross-border grain trade, specialty manufacturing B2B, professional services marketplace	3–4 types, 2–3 jurisdictions, semi-differentiated, multi-issue deals
Tier 3 — Complex	Multi-corridor agricultural trade, cross-border healthcare, defense procurement, art and collectibles	5+ types, 5+ jurisdictions, highly differentiated, complex deals

Phases 1–6: Digital Twin Prototype

Phase	Tier 1	Tier 2	Tier 3
Phase 1: Market Description	4–8 hrs	8–16 hrs	16–40 hrs
Phase 2: Knowledge Curation	8–16 hrs	24–48 hrs	60–120 hrs
Phase 3: Marketplace Configuration	4–8 hrs	12–24 hrs	24–60 hrs
Phase 4: Population Synthesis	4–8 hrs	8–16 hrs	16–40 hrs
Phase 5: Assembly & Deployment	2–4 hrs	4–8 hrs	8–16 hrs
Phase 6: Simulation & Demo	8–16 hrs	16–32 hrs	32–80 hrs
Subtotal (Phases 1–6)	**30–60 hrs**	**72–144 hrs**	**156–356 hrs**
Elapsed time (1 person)	**1–2 weeks**	**2–4 weeks**	**4–9 weeks**
Elapsed time (2 people)	**3–5 days**	**1–3 weeks**	**3–6 weeks**

Phase 7: Business Design Validation

Service integration	Tier 1	Tier 2	Tier 3
Sample fulfillment test	8–16 hrs	16–32 hrs	32–60 hrs
Market analytics prototype	8–16 hrs	16–40 hrs	40–80 hrs
Trade finance integration test	—	16–32 hrs	32–60 hrs
Compliance advisory test	4–8 hrs	8–24 hrs	24–60 hrs

Service integration	Tier 1	Tier 2	Tier 3
Phase 7 subtotal (typical scope)	**16–32 hrs**	**48–120 hrs**	**120–300 hrs**
Elapsed time	**1–2 weeks**	**3–6 weeks**	**6–12 weeks**

Phase 8: Market Launch (Technical Work)

Task	Forge effort
Clean cutover (archive synthetic, deploy)	8–16 hrs
Production infrastructure setup	16–40 hrs
Operational handoff documentation	8–24 hrs
Phase 8 technical subtotal	**32–80 hrs**

Participant recruitment and ongoing operations are the sponsor's responsibility. A trade commission launching for its existing member base can generate real activity in 3–6 months. A new entrant should plan for 6–18 months to reach minimum viable market thickness.

Full Lifecycle Summary (Tier 2)

Stage	Effort	Elapsed time	Primary responsibility
Phases 1–6: Digital Twin Prototype	72–144 hrs	1–4 weeks	MarketForge + domain expert
Phase 7: Business Design Validation	48–120 hrs	3–6 weeks	Sponsor + MarketForge
Phase 8: Market Launch (technical)	32–80 hrs	1–2 weeks	MarketForge
Phase 8: Business operations	Sponsor-funded	3–18 months	Sponsor
Total technical effort	**152–344 hrs**	**5–12 weeks**	MarketForge

The Effort Curve

Each additional vertical in the same sector costs progressively less:

- **First vertical:** 100% (baseline)
- **Second vertical (same sector):** ~65%
- **Third vertical (same sector):** ~40%
- **Fourth vertical (adjacent sector):** ~30%
- **Fifth and beyond:** ~25%

Domain schemas, prompts, and population generation patterns are reused and adapted rather than rebuilt. Cross-sector transitions reset the curve partially — but the tooling and process infrastructure

carry over.

---

7. Building the Forge — One-Time Infrastructure

Before the per-vertical workflow can operate, the integration infrastructure must be built. This is a one-time investment estimated at **16–25 weeks** of development.

Key components:

- KnowledgeSlot → Cosolvent reference library output format
- Reference library table with metadata-filtered semantic search
- Domain schema → marketplace configuration generator
- ClientSynth file-based Cosolvent export
- ClientSynth marketplace configuration import
- **Domain schema → ClientSynth vocabulary constraints** (new work)
- Deal entity and collaborative record model
- Handoff Artifact / Deal Brief generator
- Simulation runner and analytics
- End-to-end orchestration

This timeline overlaps substantially with phased work already planned in the individual project roadmaps. The Forge primarily adds three missing connectors and an orchestration layer on top of work that was already planned.

---

8. Strategic Packaging

The Forge enables three strategic models:

Sponsored First Vertical

Without the Forge, a sponsor must fund both framework development and vertical-specific work simultaneously. With the Forge, the pitch becomes:

> *"The infrastructure exists. Here is a working digital twin of your market — built from your own reference documents, populated with realistic synthetic participants, demonstrating matching, deal assembly, and Deal Brief generation. Your development budget goes toward production deployment and real participant recruitment — not R&D."*

The digital twin is the demo. The sponsor does not have to imagine what their marketplace will look like — they can click through it.

Approximate prototype costs:

Sponsor tier	Labour estimate	What the sponsor gets
Tier 1	$4,500–$9,000	Working digital twin, simple market
Tier 2	$10,800–$21,600	Working digital twin, moderate complexity
Tier 3	$23,400–$53,400	Working digital twin, complex cross-border market

Academic Partnership

The Forge transforms a "build a marketplace" research project into a **"run experiments on market physics"** research platform — a better fit for academic incentives. Graduate students build digital twins of thin markets, run controlled experiments, and publish quantitative results. A Tier 2 digital twin maps well to a single-semester graduate project.

Sample research questions the platform enables:

- "How does opacity reduction affect market thickness?"
- "What is the cold start threshold for a cross-border grain market?"
- "Do asynchronous brokerage agents reduce temporal distance effects?"
- "How does facilitator availability affect deal assembly?"

Founder / Startup Partnership

The Forge gives founders a repeatable, scalable delivery mechanism — a production line for digital twins. A potential business model: charge sponsors $15K–$50K for a Tier 2/3 prototype, then $5K–$15K/month for managed hosting. At 4–6 verticals per year, prototype revenue alone is $200K–$600K before production deployments or market intelligence revenue.

---

9. Success Criteria

The Forge is operational when all of the following are demonstrated end-to-end:

- ☐ A market description and reference documents produce a validated domain schema
- ☐ The domain schema generates a complete marketplace configuration that compiles without error
- ☐ ClientSynth generates a population of 50+ participants conforming to the marketplace schema, using domain-appropriate vocabulary
- ☐ The reference library is loaded and queryable from a participant's perspective with metadata-filtered results
- ☐ Semantic matching produces plausible buyer-seller pairs from the synthetic population
- ☐ A deal can be initiated, its parameters recorded collaboratively, facilitator roles identified, and a Deal Brief generated
- ☐ A non-technical stakeholder can walk through the prototype and understand the market it represents

---

10. Risk Register

Risk	Impact	Mitigation
Domain schema format underspecified	Produces marketplace configs that don't match real market structure	Build the first domain schema by hand for grain trade, then extract the format pattern
ClientSynth generates implausible populations	Undermines demo credibility	Build domain vocabulary constraints before attempting first vertical; human

Risk	Impact	Mitigation
		review in Phase 4
Sponsor expects production system, gets prototype	Relationship damage	Clear positioning: the digital twin is a demonstration tool. Production deployment is Phase 8.
Single-developer bottleneck	Timeline extends	All three repos are open-source and documented; a partner developer can contribute to any
Academic partner's timeline too slow	Momentum lost	Academic track runs in parallel with other models, not as sole strategy

11. Relationship to Thin Market Theory

Whitepaper concept	MarketForge implementation
Three uses of Digital Twins	Phase 6 covers all three: testing automation design, validating market physics, demonstrating viability
Cold start problem	ClientSynth-generated populations bypass the cold start entirely for prototyping
Market physics parameters	Domain schema captures market characteristics; simulation reveals which forces dominate
Trusted intermediary	Matching engine implements confidential matching; digital twin demonstrates it in action
Handoff Artifact	Deal Brief generator produces the platform's primary deliverable; Phase 7 validates it with real service providers
Framework generalization	The Forge IS the generalization — one configuration file per vertical, not hardcoded schemas
Feedback loop	Digital twin → business validation → real market → collect real data → recalibrate → iterate
Fulfillment & Settlement	Phase 7 sample fulfillment testing validates whether the Deal Brief drives real-world logistics

12. Open Questions

1. **Should the orchestration layer live in its own repository?** Arguments for: clean separation of concerns, independent release cycle. Arguments against: another repository to maintain; the orchestration is thin and mostly configuration.

2. **Should the domain schema format be published as a standard?** If the Forge produces digital twins for multiple verticals, the domain schema YAML becomes a de facto standard for describing thin markets. Is it worth formalizing as a contribution alongside the whitepaper?

3. **What is the minimum viable demo?** Phase 6 ranges from "browse and match" through "full Deal Brief assembly with facilitator coordination." What is the minimum that constitutes a convincing demo for each strategic model?

4. **What is the minimum viable Phase 7?** A plausible minimum is market analytics plus one physical service test (sample fulfillment or trade finance). What is the minimum that convinces a sponsor to commit to Phase 8?

5. **How should analytics from Phase 7 be preserved across the cutover?** Synthetic population data generates market structure insights (match density, corridor traffic, facilitator demand patterns) that may have analytical value even after synthetic users are retired. Should this be preserved as a baseline dataset?

This roadmap will be updated as implementation progresses and as the thin market framework evolves.