

Discovery Agent Design

Status: Design Complete - Ready for Implementation **Phase:** Phase 1 (Understanding) **Last Updated:** 2025-12-08

Executive Summary

The Discovery Agent is Phoenix's foundation for Phase 1 (Understanding). It conducts stakeholder interviews, builds the As-Is graph, and extracts strategic insights that enable transformation in Phase 2. Unlike traditional discovery tools that just map processes, this agent **identifies business model transformation opportunities** by asking strategic questions and connecting operational patterns to market opportunities.

Key Innovation: Adaptive conversation flow that starts strategic (business model, competitive landscape, inception story) then selectively goes deep into functions when detecting transformation signals—avoiding the trap of just automating existing processes.

Section 1: Core Purpose & Capabilities

Three Integrated Outcomes

The Discovery Agent delivers:

1. Complete As-Is Graph

- Entities: Person, Role, Team, Process, System, Document, Metric
- Relationships: REPORTS_TO, PERFORMS, USES, PRODUCES, DEPENDS_ON, BLOCKS, ENABLES
- Confidence tracking: confirmed | high | medium | low

2. Strategic Context & Narrative

- Business model and unit economics
- Competitive landscape and market position
- Inception story (how they got here)
- Leadership vision and risk tolerance
- Growth constraints and opportunities

3. Transformation-Ready Insights

- Operational bottlenecks (efficiency opportunities)
- Strategic constraints (growth blockers)
- Revenue & expansion opportunities (growth vectors)
- Business model opportunities (transformation candidates)

Success Criteria (Weighted)

- **Completeness (30%):** As-Is graph captures all critical entities and relationships
- **Insight Density (35%):** Quality and relevance of transformation insights

- **Consultant Satisfaction (20%)**: Agent accelerates work without reducing quality
- **Transformation Readiness (15%)**: Provides foundation for Vision Agent to propose bold scenarios

Graduation Threshold: Consultants report agent surfaces at least one strategic insight they hadn't yet identified.

Section 2: Adaptive Conversation Flow

Starting Point: Strategic Foundation

Every discovery begins with leadership interviews to establish:

- **Inception story**: How did this company come to exist? What problem were they solving?
- **Business model**: How do they make money? What's their unit economics?
- **Competitive landscape**: Who else does this? Why are they different/better?
- **Growth vision**: Where do they want to be in 3-5 years?
- **Constraints**: What's stopping them from getting there?

This strategic context prevents the trap of just automating existing processes without understanding if those processes should exist at all.

Detection Mechanism: When to Go Deep

The agent uses **combination detection** to decide when to drill into function-level detail:

A. Keyword Triggers

- Mentions of manual work, bottlenecks, workarounds
- "We can't scale because..."
- "Customers keep asking for..."
- References to external customers using internal tools

B. Gap Analysis

- Critical entities mentioned but not fully understood
- Contradictions between stakeholder statements
- Missing data for graph completeness

C. Consultant Guidance

- Consultant explicitly requests deep-dive on specific function
- Consultant flags area as strategically important

D. Natural Energy

- Stakeholder shows enthusiasm or frustration about specific topic
- Unprompted elaboration signals importance

Function-Level Discovery

When triggered, the agent systematically explores:

- **Roles involved:** Who does this work?
- **Current process:** Step-by-step workflow
- **Systems used:** What tools/data/integrations?
- **Pain points:** Where does it break? How often?
- **Time/cost impact:** Quantifiable bottlenecks
- **Workarounds:** What hacks exist?
- **Interdependencies:** What else breaks if this fails?

This creates the detailed As-Is understanding needed for transformation design.

Section 3: Graph Population Mechanism

Write Pattern: Real-Time with Confidence Tracking

Every entity and relationship gets a **confidence** property:

- **confirmed:** Human explicitly validated
- **high:** Strong evidence from conversation (repeated mentions, detailed description)
- **medium:** Reasonable inference (e.g., "acquisition team" implies role exists)
- **low:** Speculative connection (e.g., assumed reporting structure)

Batch Creation During Interviews

As conversations flow:

1. **Extract entities** with timestamps
2. **Create relationships** between entities
3. **Tag confidence level** based on source (direct statement vs. inference)
4. **Flag gaps** where critical information is missing

Entity Resolution

The agent handles duplication intelligently:

- **Same-session:** Maintains context to avoid re-creating entities
- **Cross-session:** Proposes merges when patterns match (e.g., "John in acquisitions" matches existing John)
- **Consultant review:** Queues ambiguous merges for human decision

Consultant Workflow

At end of each interview:

- Agent presents **inferred items** (medium/low confidence) for validation
- Consultant can confirm, reject, or add clarifying notes
- Confirmed items upgrade to high confidence
- Rejected items mark patterns to avoid repeating mistakes

This allows the graph to grow organically during discovery while maintaining accuracy through human oversight.

Section 4: Insight Extraction & Transformation Readiness

Four Types of Insights

1. Operational Bottlenecks (Efficiency opportunities)

- Pattern: High manual effort, error-prone handoffs, duplicate data entry
- Example: "15-30 min per loan creation due to BPO→Connect re-keying"
- Tags: Process inefficiency, automation candidate, quick win

2. Strategic Constraints (Growth blockers)

- Pattern: "We could grow but..." statements, capacity limitations, market opportunity gaps
- Example: "Can't scale to new cities because our process doesn't transfer without expert staff"
- Tags: Scaling constraint, platform opportunity, business model consideration

3. Revenue & Expansion Opportunities (Growth vectors)

- Pattern: Underserved markets, pricing power untapped, adjacent services requested by customers
- Example: "Customers keep asking if we offer X, but we always say no"
- Tags: Alternative revenue stream, market expansion, product extension

4. New Business Model Opportunities (Transformation candidates)

- Pattern: Unique IP, replicable methodology, potential external customers, entire new market creation
- Example: "Two other companies already pay us to use customized versions of our system"
- Tags: Platform transformation, commercialization potential, strategic pivot, new business creation

Insight Generation Process

As conversations progress:

1. **Detect patterns** across multiple statements (e.g., repeated customer requests)
2. **Quantify impact** where possible (time × frequency = annual cost OR market size × capture rate = revenue potential)
3. **Connect to strategic context** (bottleneck + growth goal = transformation opportunity)
4. **Flag for human validation** (low confidence insights marked speculative)

Transformation Readiness Signals

The agent monitors for **pivot indicators** that suggest platform transformation or new business creation:

- Existing external customers (market validation)
- Unique methodology/IP (defensible differentiation)
- Scalability constraints (current model can't support growth)
- Leadership appetite for bold change (risk tolerance)
- Repeated customer requests for adjacent services (product-market fit signals)

These signals feed directly to Vision Agent in Phase 2, providing the strategic foundation for transformation scenarios that could range from operational improvement to entirely new business models.

Section 5: Harness Integration & Consultant Interface

Harness Responsibilities

State Management:

- Conversation history persistence across sessions
- Graph write tracking (what was added when)
- Confidence level evolution (as insights get validated)
- Session context (which stakeholder, what function area)

Safety & Guardrails:

- Prevents overwriting high-confidence data with speculation
- Flags sensitive topics requiring consultant intervention (financials, personnel issues)
- Rate limits graph writes to avoid polluting with low-quality data
- Enforces validation workflows for medium/low confidence items

Observability:

- Logs all insight extractions with reasoning
- Tracks question sequences and branching decisions
- Measures conversation quality (depth, breadth, insight density)
- Provides audit trail for graph changes

Consultant Interface Patterns

During Interviews:

- Agent suggests next questions based on gaps
- Real-time confidence indicators on entities/relationships
- "Pause for clarification" prompts when detecting contradictions
- Automatic note-taking with structured extraction

Post-Interview Review:

- Summary of key findings organized by insight type
- List of inferred items requiring validation (sorted by impact)
- Gap analysis: "Still need clarity on X, Y, Z"
- Suggested stakeholders to interview next

Graph Visualization:

- As-Is graph view filtered by confidence level
- Highlight recent additions from last interview
- Show connections between insights (e.g., bottleneck → revenue opportunity)
- Flag orphaned entities (mentioned but not connected)

Section 6: Interaction with Other Platform Agents

Handoff to Vision Agent

What Discovery provides:

- Complete As-Is graph (entities, relationships, processes)
- Strategic context (business model, competitive landscape, inception story)
- Categorized insights (bottlenecks, constraints, revenue opportunities, new business models)
- Transformation readiness signals (market validation, IP uniqueness, leadership appetite)

How Vision Agent uses it:

- Reads As-Is graph to understand current state
- Analyzes insights to identify transformation vectors
- Generates multiple To-Be scenarios (efficiency play vs. platform transformation vs. new business creation)
- Uses transformation signals to calibrate boldness (conservative optimization vs. radical reimagination)

Handoff to Process Mapper Agent

What Discovery provides:

- Current process flows with handoffs mapped
- Identified bottlenecks with quantified impact
- Manual steps flagged for automation consideration
- Role-to-process relationships

How Process Mapper uses it:

- Redesigns workflows to eliminate bottlenecks
- Proposes automation opportunities with ROI estimates
- Maps To-Be processes that support transformation scenarios
- Identifies integration points between systems

Handoff to Data Architect Agent

What Discovery provides:

- Current systems and data sources
- Integration gaps (manual data re-entry points)
- Data quality issues and constraints
- Entities and relationships that need persistence

How Data Architect uses it:

- Designs unified data model for platform
- Proposes integration architecture
- Identifies data governance requirements
- Maps As-Is data to To-Be data fabric

Sentinel Integration

The Sentinel reviews Discovery outputs:

- Challenges assumptions: "Are we sure this bottleneck is real, or is it a symptom?"
 - Inverts insights: "If we eliminate this constraint, what new problems emerge?"
 - Questions incentives: "Who benefits from keeping the current process manual?"
 - Flags biases: "Are we anchoring on their stated problem instead of finding the real opportunity?"
-

Section 7: Technical Implementation Considerations

LLM Selection

Approach: Model-agnostic architecture with pluggable LLM backends

- Different models excel at different tasks (reasoning, extraction, classification)
- Cost/performance trade-offs evolve rapidly
- Clients may have preferences or compliance requirements
- Allow experimentation and optimization without rewriting agent logic

Key capabilities required:

- Strong reasoning for strategic questioning
- Excellent structured data extraction from unstructured conversation
- Context window sufficient for full interview sessions
- Function calling for graph operations

Graph Schema for Discovery

Core Entities:

- Person, Role, Team, Process, System, Document, Metric, Insight
- Each with `confidence`, `source_session`, `last_validated` properties

Relationships:

- REPORTS_TO, PERFORMS, USES, PRODUCES, DEPENDS_ON, BLOCKS, ENABLES
- Tagged with `confidence` and `inferred_from` metadata

Insight Nodes (special type):

- Type: bottleneck | constraint | revenue_opportunity | business_model_opportunity
- Impact: quantified where possible (hours/year, revenue potential, etc.)
- Evidence: links back to conversation statements
- Status: proposed | validated | rejected

Conversation Storage

Structure:

- Session-based (one session = one stakeholder interview)
- Full transcript with timestamps
- Extracted entities/relationships linked to specific utterances
- Consultant notes inline with transcript

Retrieval:

- Search across all sessions by topic/keyword
- Find all mentions of specific entity
- Trace confidence evolution over time

API Design**Discovery Agent endpoints:**

- `POST /discovery/session/start` - Initialize new interview session
 - `POST /discovery/session/message` - Send message, get agent response + graph updates
 - `GET /discovery/session/{id}/review` - Get validation items for consultant review
 - `POST /discovery/session/{id}/validate` - Approve/reject inferred items
 - `GET /discovery/insights` - Retrieve categorized insights with filtering
-

Section 8: MVP Scope & Validation with Company A

MVP Feature Set

Must Have (Week 1-4):

- Adaptive conversation flow with strategic context questioning
- Real-time graph population (entities, relationships, confidence tracking)
- Insight extraction (4 types: bottlenecks, constraints, revenue opportunities, business model opportunities)
- Post-interview validation workflow for consultants
- Basic graph visualization filtered by confidence

Should Have (Week 5-8):

- Entity resolution across sessions
- Gap analysis ("still need to learn X, Y, Z")
- Suggested next stakeholders to interview
- Insight quantification (time/cost/revenue impact)
- Integration with transcript storage

Won't Have (v1.0):

- Multi-agent coordination (Phase 2 feature)
- Workspace graph (build when Vision Agent needs it)
- Advanced analytics or predictive modeling
- Automated report generation

Validation Approach with Company A

Parallel Testing:

- Run Discovery Agent on same stakeholder interviews consultant is conducting
- Compare agent-generated insights to consultant observations

- Measure where agent adds value vs. creates noise

Success Metrics:

- **Completeness:** Does graph capture all critical entities/relationships consultant identified?
- **Insight Quality:** Do insights match or exceed consultant's findings?
- **Time Savings:** Does post-interview review reduce documentation time?
- **False Positives:** How many low-value inferences need to be rejected?

Feedback Loop:

- Weekly retrospectives with consultants on Company A
- Tune prompts based on what worked/didn't work
- Adjust confidence thresholds based on accuracy
- Refine strategic questioning based on what reveals transformation opportunities

Graduation Criteria: Discovery Agent is ready for production when consultants on Company A report it **accelerates their work without reducing quality**—specifically when it surfaces at least one strategic insight they hadn't yet identified.

Summary: Discovery Agent Design

The Discovery Agent delivers three integrated outcomes:

1. **Complete As-Is graph** with confidence tracking and validation workflow
2. **Strategic context** (business model, competitive landscape, inception story)
3. **Transformation insights** (4 types: bottlenecks, constraints, revenue opportunities, business model opportunities)

Key Design Principles:

- Start strategic, go deep selectively (avoid just mapping existing processes)
- Real-time graph population with human validation
- Model-agnostic, API-driven architecture
- Validated against real consultant work (Company A engagement)

This agent becomes the foundation of Phase 1 (Understanding), enabling bold reimagination in Phase 2 by ensuring we deeply understand the As-Is before proposing the To-Be.

Next Steps:

1. Implement harness infrastructure (state management, safety, observability)
 2. Build graph schema and API endpoints
 3. Develop adaptive conversation prompts based on Company A learnings
 4. Create consultant validation interface
 5. Run parallel testing with Company A engagement
-

References

- [Company A Learnings](#) - Real-world validation case

- [Phoenix Mental Model](#) - Overall platform architecture
- [The Sentinel Agent](#) - Critical validation perspective