# Cartedo Content Flow Analysis

## Based on Shweta's Requirements vs Current Implementation

**Date:** 2026-01-16 **Purpose:** Deep analysis of content flow expectations vs actual implementation **Source:** Slack conversations (Dec 18, Dec 22, Jan 9)

---

## Part 1: Shweta's Three-Step Framework

### What Shweta Asked For

| Step | Requirement | Expected Output |
|---|---|---|
| **1. Glossary** | Simple table of agents with one-line descriptions | Agent name, task description, threshold value, why |
| **2. KPI Dashboard** | Metrics to track progress | Pass/fail columns, reasons, suggestions |
| **3. 20 Runs** | Same simulation, 20 different scenarios | "19 on 20" style with failure reasons |

### Shweta's Exact Words (Jan 9 Meeting)

> "First give me a glossary of what you're doing. What are the agents? Within each agent, what are the tasks? What is your threshold value for that, and why?"
>
> "I want you to create a dashboard of those agent tasks as KPIs. How do I know I'm making progress?"
>
> "You have done 20 different simulation content, and you have to recontextualize it... 19 on 20 and the reason why."

---

## Part 2: Current System vs Shweta's Expectations

### Gap Analysis

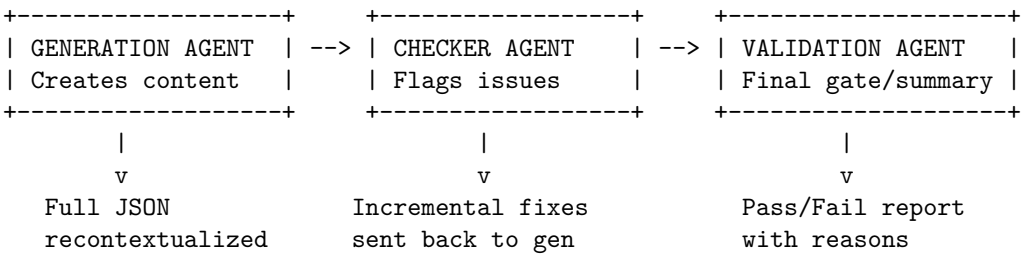
| Shweta's Expectation | Current Implementation | Gap |
|---|---|---|
| Simple glossary in plain English | Complex code documentation | NEED: Readable agent summary |
| 98% starting threshold | Mixed thresholds (85%-100%) | NEED: Standardize at 98% |
| Dashboard for quick decisions | JSON output files | NEED: Visual dashboard |
| Domain fidelity checking | Poison term checking only | NEED: Industry KPI validation |

| Shweta's Expectation | Current Implementation | Gap |
|---|---|---|
| Context fidelity checking | KLO alignment checking | PARTIAL: Needs strengthening |
| Resource quality checking | Basic content checks | NEED: Answer leakage detection |
| 20 different scenarios tested | 4 runs (FAILED14-20) | NEED: More scenario variety |
| Non-technical format | Technical JSON/logs | NEED: Executive summary |

---

## Part 3: Content Flow - What Shweta Expects

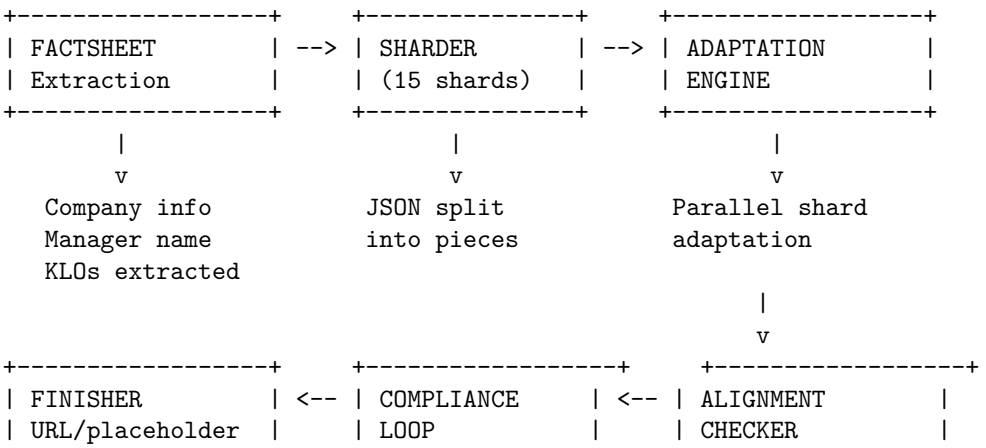
### The Ideal Pipeline (From Dec 22 Meeting)




### Shweta's Agent Requirements (Exact Quote)

> "You will have a generation agent. Then you will have a checker agent. Then you will have a validation agent that the checker agent basically sends only the incremental changes that the generation agent needs to regenerate."

### Current Pipeline (What We Built)



```
| cleanup          |     | Max 5 retries  |     | 9 rules checked  |
+------------------+     +----------------+     +------------------+
```

---

## Part 4: Deep Analysis - Three Core Fidelities

**Shweta's Three Fidelities (Dec 22 Meeting)**

**1. Domain Fidelity**

> "Domain fidelity means that the domain of the industry... In fast food, it's a $1 menu or 50% off on a burger. In airlines, you have upgrade to business class, loyalty points. In telcos, it is ARPU, churn rates."

**What She Wants:** - Industry-specific terminology correctly used - KPIs appropriate to the new domain - No old domain language leaking through

**Current Implementation:** - Poison term checking (removes old entity names) - No explicit industry KPI validation - HR terminology still leaking (185 occurrences in FAILED20)

**Gap:** System checks for poison terms but doesn't validate that NEW industry terminology is correctly applied.

---

**2. Context Fidelity**

> "If the goal was go/no-go decision and choosing between 4 strategic options, are you still doing that? That was the main goal of the topic."

**What She Wants:** - Original learning objectives preserved - Same pedagogical structure maintained - Simulation flow unchanged

**Current Implementation:** - KLO-to-Questions alignment (90%) - KLO-to-Resources alignment (88%) - Scenario coherence checking (90%)

**Gap:** Context is partially preserved but alignment scores are below 95% target.

---

**3. Resource Quality**

> "Does the resource contain all the information the student needs to answer the submission questions? Is it self-contained? Is it within 1500 words? Does it NOT have the answer?"

**What She Wants:** - Self-contained resources - Under 1500 words - Enables inference, doesn't give answers - "Dots to connect, not connected dots"

**Current Implementation:** - Resource completeness checking (2/4 runs pass) - No explicit word count enforcement - No answer leakage detection - Content truncation issues (PESTEL analysis cut off)

**Gap:** Major gap in ensuring resources don't reveal answers directly. This is a critical pedagogical requirement.

---

## Part 5: The Inference Map Problem

**Shweta's Key Insight (Dec 22 Meeting)**

> "You need to do an inference map first. The resource should be self-contained, but it should not carry the answer. It should basically have all the dots for inference to connect the dots, but it doesn't really give the connected dots to the student."

### What This Means

| Correct Approach | Wrong Approach |
|---|---|
| Provide market size data | Say "the TAM is $50B" directly |
| Show competitor pricing | Conclude "our prices should be X" |
| Present PESTEL factors | Provide the final analysis |
| Give financial inputs | Calculate the break-even point |

### Current Status

**Problem:** When temperature is set to 1.0 and prompt says "self-contained", LLMs tend to include answers to make resources complete.

**Evidence from FAILED20:** - KLO-to-Resources alignment at 88% - Resources described as "truncated mid-PESTEL" - No explicit inference map validation

**Required Fix:** 1. Add "inference map" agent that checks resources don't contain answers 2. Validate that submission questions can be answered FROM resources but not BY resources 3. Add explicit rule: "Resource provides data points. Student provides analysis."

---

## Part 6: Threshold Philosophy

### Shweta's Approach (Dec 22 Meeting)

> "You start with 98%, and if you are not getting it, you come down to 95%... After you've got me to 98, 99% accuracy, let's look at the 2% where we are losing."

### Current Threshold Configuration

| Check Type | Shweta's Target | Current Setting |
|---|---|---|
| Domain Fidelity | 98% | Not explicitly set |
| Context Fidelity | 98% | 95% |
| Resource Quality | 98% | Not explicitly set |
| Critical Checks | 100% | 100% |
| Alignment Rules | 95% | 95% |
| Flagged Checks | 85-90% | 85-90% |

### Gap

System should START at 98% and only reduce if justified. Current system starts at 95% or lower.

---

## Part 7: Reporting Format

### What Shweta Wants

> "I'm not going to review your JSON. I want you to give me reports and dashboards."
>
> "I don't want essays. I don't have the time. Make it into a simple table."
>
> "I will not read beyond 3 pages, so you have to keep it very clean and very high level."

**Ideal Report Format (From Shweta)**

| Agent | Task | Threshold | Run 1 | Run 2 | Run 3 | ... | Reason for Fail |
|---|---|---|---|---|---|---|---|
| Domain Checker | Verify industry terms | 98% | PASS | PASS | FAIL | ... | "Airlines" appeared |
| Context Checker | KLO preservation | 98% | PASS | FAIL | PASS | ... | KLO3 not assessed |
| Resource Checker | Self-contained, no answer | 98% | FAIL | FAIL | PASS | ... | Answer in resource |

**Current Report Format**

- JSON files with detailed logs
- VALIDATION_REPORT.md created (better but still complex)
- VALIDATION_DASHBOARD.md created (closer to expectation)

---

## Part 8: Ken's Five Simulations

**Context (Dec 22 Meeting)**

> "Currently we have 5 simulations for a professor called Ken. Ken is a management and strategy professor who has done 5 simulations with us. Now he wants to repeat those simulations, but he wants to change the scenario."

**Target Deliverable**

| Simulation | Original Scenario | Example New Scenario |
|---|---|---|
| Sample 1 | Gen Z organic T-shirts | Pet food market entry |
| Sample 2 | Fast food $1 menu response | Airline loyalty program |
| Sample 3 | Functional beverage launch | SaaS product-market fit |
| Sample 4 | TBD | TBD |
| Sample 5 | TBD | TBD |

**Current Status**

Only Sample 1 (HR Hiring -> EcoChic Threads) has been tested across multiple runs. Need to: 1. Get remaining 4 simulation PDFs 2. Test each across 20 different scenarios 3. Build confidence across all 5 Ken simulations

---

## Part 9: Critical Questions for Seniors

Based on this deep analysis, here are questions that reveal system gaps:

**Domain Fidelity Questions**

1. **Q:** How do we validate that industry-specific KPIs are correctly applied?
   - Current: We check for poison terms (old names)
   - Missing: Validation that NEW industry terms are present and correct
   - Example: If adapting to telco, do we verify ARPU, churn rate, ARPM appear?
2. **Q:** Where is the industry KPI lookup table?
   - Shweta mentioned fast food ($1 menu), airlines (loyalty points), telcos (ARPU)
   - System should have reference data for each industry's expected terminology

**Context Fidelity Questions**

3. **Q:** How do we verify the original learning goal is preserved?
   - Example: "Go/no-go decision with 4 strategic options"
   - Is there a check that these 4 options still exist after adaptation?
4. **Q:** What happens when KLOs conflict with new scenario?
   - HR Hiring KLOs don't cleanly map to Market Entry
   - Who decides what KLO modifications are acceptable?

**Resource Quality Questions**

5. **Q:** How do we detect if resources contain answers?
   - Shweta's "inference map" requirement is not implemented
   - What agent checks that students must connect dots, not read conclusions?
6. **Q:** Why is there no word count enforcement?
   - Shweta specified "under 1500 words"
   - Current resources are being truncated but not length-validated

**System Architecture Questions**

7. **Q:** Why do we have 15 shards instead of Shweta's 3-agent model?
   - Shweta: Generation -> Checker -> Validation
   - Current: Factsheet -> Sharder -> Adaptation -> Alignment -> Compliance -> Finisher
   - Is the added complexity justified?
8. **Q:** Why doesn't the checker agent send "incremental changes" back?
   - Shweta specifically asked for this pattern
   - Current: Alignment Fixer runs full fixes, not incremental

**Reporting Questions**

9. **Q:** Why are we still generating JSON instead of dashboards?
   - 3 months in, still no simple table output
   - VALIDATION_DASHBOARD.md is manual, not auto-generated
10. **Q:** What is blocking 98% threshold achievement?
    - Currently stuck at 92.33% alignment
    - Specific blockers: HR terminology (185), KLO gaps (88-90%)
    - Clear path to fix exists but not implemented

---

## Part 10: Recommendations

**Priority 1: Implement Inference Map Checking**

```
# Pseudo-code for inference map agent
def check_inference_map(resource, submission_questions):
    """
```

```
    Verify resource provides DATA but not ANSWERS
    """
    for question in submission_questions:
        # Check if resource contains data needed to answer
        data_present = contains_relevant_data(resource, question)

        # Check if resource DOESN'T contain the answer directly
        answer_absent = not contains_direct_answer(resource, question)

        if not (data_present and answer_absent):
            return FAIL, f"Question {question.id}: Data={data_present}, Answer leaked={not answer_absen

    return PASS, "Inference map valid"
```

**Priority 2: Add Industry KPI Validation**

Create lookup table:

```
fast_food: ["$1 menu", "combo deals", "drive-through", "franchise"]
airlines: ["loyalty points", "seat upgrade", "fare class", "load factor"]
telco: ["ARPU", "churn rate", "subscriber", "data bundle"]
market_entry: ["TAM", "SAM", "SOM", "CAC", "LTV", "break-even"]
```

**Priority 3: Auto-Generate Dashboard**

Convert current JSON outputs to simple tables automatically after each run.

**Priority 4: Test All 5 Ken Simulations**

Request remaining 2 PDFs from Shweta and build test suite across all 5.

---

## Part 11: Timeline Mismatch

**What Was Promised**

| Date | Commitment | Actual |
|---|---|---|
| Dec 29 | "Generation down to 40-50 seconds" | Achieved but accuracy dropped |
| Jan 5 | "Demo 5 simulations working" | Missed - only 1 simulation tested |
| Jan 8 | "Stable version by Friday" | Missed - still at 92% alignment |
| Jan 9 | "Glossary + Dashboard by Monday" | Partial - VALIDATION_DASHBOARD created |

**Root Cause**

> "In the final phase, I attempted to optimize latency using chunked generation and later RAG. That optimization removed some implicit global constraints that were holding alignment together across sections." - Poovendhan (Jan 9)

**Lesson Learned**

Shweta's advice was clear: > "First ensure accuracy. Don't worry about time and tokens just yet. After you've got me to 98%, then look at optimization."

## Part 12: Success Criteria Summary

**From Shweta's Perspective**

| Criteria | Measure | Target |
|---|---|---|
| Domain Fidelity | Industry terms correctly applied | 98% |
| Context Fidelity | Learning objectives preserved | 98% |
| Resource Quality | Self-contained, no answers | 98% |
| Speed | Per simulation | $< 5$ minutes |
| Coverage | Ken's simulations | 5/5 working |
| Scenarios | Different contexts | 20 per simulation |
| Report Format | Simple tables, 3 pages max | Achieved |

**Current Status**

| Criteria | Current | Gap |
|---|---|---|
| Domain Fidelity | ~85% (HR terms leaking) | -13% |
| Context Fidelity | 90% (KLO alignment) | -8% |
| Resource Quality | 88% (truncation issues) | -10% |
| Speed | ~2 minutes | ACHIEVED |
| Coverage | 1/5 simulations | -4 simulations |
| Scenarios | 4 runs | -16 scenarios |
| Report Format | DASHBOARD.md created | ACHIEVED |

## Conclusion

The current system is approximately **70% aligned** with Shweta's expectations:

**What's Working:** - Speed is acceptable - Report format is improving - Poison term removal (no old entity names) - Compliance loop working (100% structural integrity)

**What's Missing:** - Industry KPI validation (domain fidelity) - Inference map checking (resource quality) - 98% threshold enforcement - 20 scenario test coverage - Remaining 4 Ken simulations

**Key Insight:** The system optimized for speed before accuracy was achieved. Shweta explicitly warned against this: "First accuracy, then speed." The current 92.33% alignment needs to reach 98% before any further optimization.

*Analysis based on Slack transcripts from Dec 18, Dec 22, and Jan 9 meetings. Generated: 2026-01-16*