

SAMPLE · FILLED EXAMPLE · ACMECORP INSURANCE

The Data Readiness Audit — Sample

AcmeCorp Insurance · Claims-Triage Agent · May 2026

The Five Dimensions of Data Readiness

Data audited: Guidewire ClaimCenter · Policy DB · Claimant history · Routing queue · Regulatory constraint store

ACCESSIBILITY

Can the agent reach the data it needs, in the format it needs, at the speed it needs?

AcmeCorp: Guidewire REST API v2.1 is stable and documented. Policy DB has a read-only API. p95 query time: 1.1 sec.

QUALITY

Is the data accurate, complete, consistent, and free of the errors that would mislead the agent?

AcmeCorp: Completeness 97.3% (above 95% threshold). Duplicate rate 0.8%. Ground-truth validation: Jan 2026.

LINEAGE

Do you know where each data element came from, how it was transformed, and who changed it last?

AcmeCorp: Lineage captured in Collibra for 80% of fields. Remaining 20% (legacy batch jobs) is a known gap.

PERMISSIONS

Is access to data governed at the field level, with the agent having only what it needs?

AcmeCorp: Agent scoped to 14 of 340 available fields. Permissions reviewed April 2026. Access logged to SIEM.

FRESHNESS

Is the data current enough for the decisions the agent will make?

ACMECORP RESULT: Overall score 31/40 (78%) → **READY TO PROCEED** with one remediation item (lineage Remediation: Complete Collibra lineage for legacy batch jobs by June 30, 2026. Owner: Data Engineering team.

Audit Questions — AcmeCorp Completed Assessment

Auditor: Priya Sharma, Data Governance Lead · Date: May 5, 2026 · ● = selected · Notes show evidence

ACCESSIBILITY	QUESTION	YES	PART	NO	EVIDENCE / NOTES
Q1	Is the data accessible via a documented, stable API or query interface?	YES			Guidewire REST API v2.1. Docs at internal wiki. SLA: 99.9% uptime (met last 12 months).
Q2	Can the agent retrieve data in under 2 seconds for 95% of queries?	YES			p95 query time: 1.1 sec (Guidewire), 0.8 sec (Policy DB). Both well within 2-sec target.
Q3	Is the data available in a machine-readable format (JSON, SQL, vector) without manual transformation?	YES			All sources return JSON. No ETL required before passing to agent.
Q4	Is there a fallback if the primary data source is unavailable?		PART		Guidewire fallback: read-only replica. Policy DB fallback: not yet configured. Q3 backlog.
ACCESSIBILITY SCORE					7/8

QUALITY	QUESTION	YES	PART	NO	EVIDENCE / NOTES
Q1	Is data completeness monitored and above 95% for fields the agent will use?	YES			Completeness: 97.3% across 14 agent fields. Monitored via Great Expectations daily.
Q2	Are duplicate records identified and resolved on a defined cadence?	YES			Dedup job runs nightly. Current duplicate rate: 0.8% (target <1%). Last run: May 4.
Q3	Is there a data quality SLA that the data team is accountable to?		PART		Informal SLA exists (>95% completeness). Formal SLA with accountability not yet signed.
Q4	Has the data been validated against a ground-truth sample in the last 90 days?	YES			Ground-truth validation: Jan 15, 2026 (110 days ago). Next scheduled: June 2026.
QUALITY SCORE					7/8

LINEAGE	QUESTION	YES	PART	NO	EVIDENCE / NOTES
Q1	Can you trace any data element back to its source system and transformation history?		PART		80% of fields have Collibra lineage. 20% (legacy batch) have manual docs only.
Q2	Is lineage captured automatically (not manually documented)?		PART		Collibra captures lineage for Guidewire and Policy DB. Legacy batch jobs: manual only.
Q3	Can you reconstruct the exact data state the agent saw at the time of any given decision?	YES			Event sourcing in ClaimCenter allows point-in-time reconstruction. Tested successfully.
Q4	Is lineage accessible to auditors without requiring engineering support?		PART		Collibra portal available to auditors. Legacy lineage requires engineering to pull.
LINEAGE SCORE					4/8

PERMISSIONS	QUESTION	YES	PART	NO	EVIDENCE / NOTES
Q1	Is the agent's data access scoped to only the fields and records it needs?	YES			Agent reads 14 of 340 fields. Scoped by service account with field-level permissions.
Q2	Are permissions reviewed and recertified at least quarterly?	YES			Last recertification: April 14, 2026. Next: July 14. Tracked in ServiceNow.
Q3	Is there a deny-list of data categories the agent must never access?	YES			Deny-list: financial reserves, SSN full, attorney notes, internal fraud flags.
Q4	Is all data access logged with the agent's identity and timestamp?	YES			All access logged to Splunk SIEM with agent service account ID and timestamp.
PERMISSIONS SCORE					8/8

FRESHNESS	QUESTION	YES	PART	NO	EVIDENCE / NOTES
Q1	Is the maximum acceptable data age defined for each data source the agent uses?		PART		Policy data: 15-min refresh (defined). Claim history: real-time (defined). Regulatory: no SLA defined.
Q2	Is data freshness monitored with alerts when SLAs are breached?		PART		Policy and claim freshness monitored. Regulatory constraint store: no monitoring yet.
Q3	Does the agent know the timestamp of the data it is acting on?	YES			All API responses include `data_as_of` timestamp. Agent logs this with each decision.
Q4	Is there a process to pause agent actions when data freshness falls below threshold?			NO	No automated pause mechanism. Manual process only. Automated circuit breaker on Q3 roadmap.
FRESHNESS SCORE					5/8

TOTAL DATA READINESS SCORE: 31 / 40 = 78% → **READY TO PROCEED**
 78% ≥ 75% threshold. Remediation item: complete lineage for legacy batch jobs (owner: Data Eng, due June 30).