

TECH & AI CASE STUDY - PHARMA - 2026

Data Catalog and Governance That Cut Audit Prep 70%

Based on a delivery engagement with Corvale Pharmaceuticals

-70%

AUDIT PREP EFFORT

-64%

SEARCH-TO-ANSWER TIME

310

CERTIFIED DATA ASSETS

0

DATA-INTEGRITY FINDINGS

A fillable practice edition: the full story, 8 actions,
6 reflections and a 90-day plan - work it with your team.

Free - no email required
emersontechnology.com

MAKE THIS PRACTICE BOOK YOURS

Work it, do not just read it

This practice book turns the Corvale Pharmaceuticals case study into a working session for your team. Read the full story first - what happened, what was built and what it changed. Then tick the 8 actions you can commit to, write short answers to the 6 reflections, and finish with a 90-day plan. Every field in this PDF is fillable: open it in Adobe Acrobat Reader, Chrome, Edge or macOS Preview, type directly into the boxes and save.

YOUR NAME	ROLE
COMPANY	TEAM / DEPARTMENT
DATE	EMAIL (OPTIONAL)

MY GOAL FOR THIS PRACTICE SESSION

- 01 The story** - What happened at Corvale Pharmaceuticals - before, during and after the project.
- 02 The results** - The numbers that changed, plus industry benchmarks from public research.
- 03 The 8-action checklist** - Tick boxes as your organization completes each step. Owner and date fields included.
- 04 The 6 reflections** - Fillable questions that pressure-test your own roadmap.
- 05 My 90-day plan** - Three fillable phase cards that turn this case study into your roadmap.

RUN IT AS A 90-MINUTE TEAM SESSION

25 min READ Read the story together. Ask everyone to note the one number that surprised them most.	40 min PRACTICE Work the 8-action checklist and the 6 reflections. Assign an owner and a date to every action you commit to.	25 min COMMIT Fill in the 90-day plan. One goal per phase, one number that tells you it worked. Book a review date before you leave the room.
--	--	---

THE STORY

What happened, and why it worked

01 The Company

Corvale Pharmaceuticals is a specialty pharma company of roughly 900 employees that develops and manufactures prescription therapies, with a portfolio spanning established molecules and a pipeline of niche indications. The business sits squarely inside the regulated world: batch records, stability studies, deviations and quality documentation are subject to FDA oversight, and under 21 CFR Part 11 the company must maintain complete, auditable electronic records. In this industry, data lineage is not an analytics nicety — it is a compliance obligation that inspectors can and do probe.

Like many mid-size pharma companies, Corvale had grown by acquisition. Two acquired companies brought their own warehouses and reporting conventions, layering on top of an ERP, laboratory information systems and a persistent universe of clinical spreadsheets. The science was excellent. The data plumbing beneath it was a liability.

02 Where Things Stood

Critical data lived everywhere and belonged to no one. A batch record table might sit in the ERP, feed a quality report, and be referenced by a spreadsheet on a network drive — and nobody could authoritatively answer what it contained, how it was derived or who owned it. Answering that question took days of archaeology: an analyst would trace table references through old ETL jobs, message former employees, and reconstruct history from memory. Every report began with the same excavation, and every audit request multiplied it.

Analysts duplicated work constantly because there was no way to know which dataset was certified and which was a half-finished experiment. Two people preparing the same metrics from different tables would produce different numbers, and resolving the discrepancy took longer than building both reports had. McKinsey's research estimates that knowledge workers spend nearly 20% of the workweek — about 1.8 hours a day — searching for and gathering information; Corvale's analysts and subject-matter experts were spending multiples of that, because the search itself required knowing who to ask.

The sharpest pain arrived with regulators. When the quality unit or an inspector requested documentation — lineage for a critical table, definitions behind a reported figure, evidence of access controls — subject-matter experts spent three to four weeks assembling it by hand: screenshots, exported table definitions, emailed attestations, reconstructed pipeline diagrams. The last audit surfaced two data-integrity findings, and while they were closeable, leadership understood the message. Gartner estimates that poor data quality costs the average organization \$12.9 million per year, much of it invisible in departmental budgets; at Corvale, the cost was concentrated in inspection risk and in the scarcest resource the company had — expert time that should have gone to science.

03 The Challenge

The challenge was that Corvale's governance existed in people's heads rather than in systems. Ownership, definitions and lineage were real but unrecorded, so every departure, reorganization or acquisition silently erased part of the institutional memory. Documentation was written for the moment — assembled for an inspection, then left to rot — which meant it was accurate when created and unreliable a year later. The company did not need more documents; it needed evidence that maintained itself.

There was also a sequencing problem. A catalog tool could be purchased in weeks, but a catalog full of unowned, unclassified, uncategorized assets would be another abandoned library. Any program had to make the catalog the

place where answers actually live: who owns this table, what does this term mean, which dataset is trusted, what changed last month — and it had to do so in a way that satisfied GxP expectations for control and traceability. Leadership set a blunt success criterion: the next inspection should be answered from the platform in days, not from archaeology in weeks.

04 Discovery and Diagnosis

Emerson Technology opened with a three-week discovery across Corvale's data landscape. The team inventoried critical data assets across the ERP, laboratory systems, the two acquired warehouses and the spreadsheet estate, and mapped which business processes depended on which tables. Ownership interviews followed: for the forty most critical tables, the team asked who owned the data, who fixed it when it was wrong, and who had approved its last structural change. For a third of those tables, the honest answer to all three questions was "nobody, currently."

The diagnosis confirmed the pattern behind the audit pain. Lineage existed only as tribal knowledge, so every evidence request restarted from zero. Classification was absent — nobody distinguished GxP-critical tables from ordinary reporting tables, so protective effort was spread thin instead of concentrated where regulators would look. Access reviews were manual and slow, and change history was reconstructable only from deployment logs and memory. The audit's two data-integrity findings traced directly to these gaps: documentation that no longer matched the system, and access evidence that could not be produced completely. The program's shape fell out of the diagnosis — name owners, harvest lineage automatically, classify against real obligations, and make evidence export a platform feature rather than a fire drill.

05 The Solution

Emerson Technology deployed an enterprise data catalog as the single entry point for data discovery and governance, and wired it into the systems where Corvale's data actually lives. The design has three layers, each aimed at one of the diagnosed failures.

Automated lineage as the foundation

The catalog is populated with automated lineage harvested from Snowflake query history, Airflow DAGs and ETL configurations — not from interviews and whiteboard diagrams. When a batch record table feeds a quality report, the path between them is visible and current, because it is read from the systems themselves. Automated lineage stays accurate as pipelines evolve; hand-drawn lineage is outdated the day it arrives. For the acquired companies' warehouses, lineage harvesting also revealed redundant and orphaned tables, giving the team a defensible basis for archiving rather than carrying every legacy asset forward.

Ownership, classification and enforced policy

Every critical data asset received a named owner and steward, recorded in the catalog with review dates, plus a data classification drawn from Corvale's actual regulatory obligations: GxP-critical, PII and confidential. Classifications are not decorative labels — they bind to role-based access controls and row-level security in the platform, so policies apply automatically at query time. A glossary of 2,400 business terms, built in workshops with quality, regulatory and commercial teams rather than by IT alone, maps technical assets to the language those teams actually use, so a quality associate can find the stability dataset without knowing its schema name. Certification workflows mark approved datasets with a visible verified badge, ending the guesswork about which copy of the truth to use.

Audit evidence on demand

For inspection responses, the platform generates evidence where it used to be manufactured: lineage diagrams, access history and change logs are exported per asset in hours, and every change to a certified asset is logged automatically for inspection. No screenshots, no reconstructing history from deployment notes, no emailing five people for attestations. When a regulator asks how a reported figure was derived and who touched the underlying pipeline since last year, the answer is a governed export rather than a three-week project. This layer converted compliance from a recurring emergency into a query.

06 The First 90 Days

Weeks one and two completed the critical-asset inventory and the ownership workshops. Assigning owners was deliberately treated as an organizational decision, not a database update: the quality unit co-signed the GxP-critical designations, and every named steward received a short charter describing what they approve and what they escalate. Weeks three through six stood up the catalog, harvested lineage from Snowflake, Airflow and the ETL layer, and ran the first glossary workshops with quality, regulatory and commercial teams — 2,400 terms would eventually land, but the first hundred covered the vocabulary of batch records, stability and deviations, which is where inspection questions concentrate.

Weeks seven through ten ran certification in waves. The forty most critical tables were certified first, each with an owner, a steward, a classification, documented quality checks and a verified badge; certification expanded outward to 310 assets as teams nominated their own datasets and met the same bar. Access policies were bound to classifications in the same window, with row-level security restricting sensitive records by role and site. Weeks eleven and twelve closed the quarter with the exercise leadership had asked for: a timed mock inspection in which the quality unit issued a realistic regulator's data request and the team responded end to end from the platform. The first run produced evidence packs in days rather than weeks, surfaced a handful of lineage gaps while they were still cheap to fix, and gave the SMEs their first experience of an inspection that did not consume a month.

07 The Results

Across two inspection cycles, audit preparation effort fell 70% — from three to four weeks of scarce subject-matter-expert time down to under a week of guided export and review. Analyst search-to-answer time dropped 64%, because certified datasets are discoverable in minutes rather than reconstructed through archaeology; the McKinsey benchmark on information search now describes a problem Corvale has largely engineered away for its data workflows. The 310 certified data assets carry named owners and automated lineage, and the most recent quality review closed with zero data-integrity findings. Regulators received evidence packs in days rather than weeks, and the SME hours previously consumed by documentation now go to science.

METRIC	BEFORE	AFTER
Audit prep effort	3-4 weeks	Under 1 week (-70%)
Search-to-answer time	-	-64%
Certified data assets	-	310
Data-integrity findings	2	0

The cultural shift matters as much as the numbers. "Who owns this table" has an answer, recorded and current. Analysts start from certified datasets instead of rebuilding the same extract for the fourth time, and disagreements about numbers now resolve by inspecting lineage rather than debating seniority. Compliance documentation is true this week, not just true when it was written — which is, in the end, exactly what an inspector is testing for.

08 What Your Business Should Do

The pattern behind Corvale's 70% reduction is available to any regulated organization, and it starts before any tool is chosen. Name an owner for every critical asset — governance begins with accountability, not with a platform purchase, and unowned data cannot be governed no matter what software sits on top of it. Harvest lineage automatically from query history and pipeline configuration rather than interviewing engineers for diagrams; automated lineage stays current, and hand-drawn lineage is outdated the day it arrives. Classify data against your real obligations — GxP-critical, PII, confidential, or whatever your regulator's vocabulary is — because generic classification schemes waste effort and miss the assets that matter most.

Build the glossary with the business, not for it: run the workshops with quality, regulatory and commercial teams, because terms defined by IT alone get ignored by the people they were meant to serve. Certify datasets with visible badges so analysts trust and reuse them — certification is what gives the business a reason to come to the catalog at all. Enforce policy in the platform by binding classifications to role-based and row-level access controls, since governance documents nobody enforces are just decorations for the next audit. Automate audit evidence export so lineage, access history and change logs are one click per asset; every hour of preparation you automate is an hour of expert time returned to real work. And run a mock inspection annually, timed end to end, to expose gaps while they are cheap and keep the documentation muscle alive.

Then ask the questions Corvale had to answer the hard way: if a regulator requested lineage for a critical table tomorrow, how many days would your team need; who owns your most important datasets — can you name the person without looking; how much of your compliance documentation is true this week versus true when it was written. Gartner's estimate that poor data quality costs the average organization \$12.9 million a year is not an abstraction in a regulated industry — it shows up as inspection findings, delayed submissions and burned-out experts. Work through the eight actions above with your quality and data leaders jointly, and revisit the reflection questions in this practice book each quarter so the governance you build stays lived rather than laminated.

REAL-WORLD PROOF & BENCHMARKS

McKinsey research estimates that knowledge workers spend nearly 20% of the workweek — about 1.8 hours a day — searching for and gathering information. - McKinsey & Company

Gartner estimates that poor data quality costs the average organization \$12.9 million per year, much of it invisible in departmental budgets. - Gartner, 2021

FDA regulations under 21 CFR Part 11 require pharmaceutical companies to maintain complete, auditable electronic records — making data lineage a compliance obligation, not a luxury. - FDA, 21 CFR Part 11

THE PRACTICE PAGES

The 8-action checklist

Work through the list with your team. Tick each box as your organization completes the step, and fill in an owner and a target date so every action has a name and a deadline.

☐

1. Name an owner for every asset

Governance starts with accountability, not tooling. Assign every critical dataset a named owner and steward before any platform decision is made — unowned data cannot be governed.

OWNER

TARGET DATE

STATUS

--	--	--

☐

2. Harvest lineage automatically

Pull lineage from query history and pipeline configuration rather than interviewing engineers for diagrams. Automated lineage stays current; hand-drawn lineage is outdated the day it arrives.

OWNER

TARGET DATE

STATUS

--	--	--

☐

3. Classify data against real obligations

Tag assets as GxP-critical, PII or confidential based on your actual regulatory obligations. Generic classification schemes waste effort and miss the assets that matter most.

OWNER

TARGET DATE

STATUS

--	--	--

☐

4. Build the glossary with the business

Run glossary workshops with quality, regulatory and commercial teams. Terms defined by IT alone get ignored by the people the definitions were meant to serve.

OWNER

TARGET DATE

STATUS

--	--	--

☐

5. Certify datasets with visible badges

Mark approved datasets clearly so analysts trust and reuse them. Certification gives the business a reason to come to the catalog in the first place.

OWNER

TARGET DATE

STATUS

--	--	--

☐ **6. Automate audit evidence export**
Configure one-click export of lineage, access history and change logs per asset. Every hour of preparation you automate is an hour of SME time returned to science.

OWNER

TARGET DATE

STATUS

☐ **7. Enforce policy in the platform**
Bind classifications to role-based and row-level access controls so policies apply automatically at query time. Governance documents nobody enforces are just decorations for the next audit.

OWNER

TARGET DATE

STATUS

☐ **8. Run a mock inspection annually**
Rehearse a regulator's data request end to end every year, timed. The mock exposes gaps while they are still cheap and keeps your documentation muscle alive.

OWNER

TARGET DATE

STATUS

AFTER THE CHECKLIST: OUR TOP 3 PRIORITIES

From the eight actions above, write down the three you will start first - and the single number you expect each one to move. Three priorities with owners beat eight without any.

PRIORITY 1 NUMBER IT SHOULD MOVE	PRIORITY 2 NUMBER IT SHOULD MOVE	PRIORITY 3 NUMBER IT SHOULD MOVE
--	--	--

THE PRACTICE PAGES

Reflections for your team

Use these questions in your next planning session and type short answers into the boxes. The gaps you find are usually the fastest wins.

Q1. If a regulator requested lineage for a critical table tomorrow, how many days would your team need?

Q2. Who owns your most important datasets — can you name the person without looking?

Q3. How many hours did your last audit consume from your scarcest subject-matter experts?

Q4. When analysts need data, do they search a catalog or message the one person who knows?

Q5. What data-integrity findings would a surprise inspection most likely surface in your environment today?

Q6. How much of your compliance documentation is true this week, versus true when it was written?

THE PRACTICE PAGES

My 90-day plan

Finish the session by committing to a plan. Three phases, one goal each, one number that tells you it worked. Bring this page to your next leadership review - or to a free consultation with Emerson Technology, and we will pressure-test it with you.

Days 1-30 Focus: audit and alignment

PRIMARY GOAL

KEY ACTIONS

HOW WE MEASURE SUCCESS

OWNER

Days 31-60 Focus: build and prove

PRIMARY GOAL

KEY ACTIONS

HOW WE MEASURE SUCCESS

OWNER

Days 61-90 Focus: scale and measure

PRIMARY GOAL

KEY ACTIONS

HOW WE MEASURE SUCCESS

OWNER

Tip: save a copy per team so each group keeps its own answers. Revisit the reflections every quarter - the questions stay the same even as the numbers move.

BOOK THE REVIEW BEFORE YOU CLOSE THE BOOK

Plans without a review date drift. Before you close this practice book, agree on when you will come back to these pages, who will run that session, and when you will check the numbers again. If you want an outside pair of eyes, bring this page to a free consultation with Emerson Technology and we will pressure-test the plan with you.

REVIEW DATE	REVIEWED BY	NEXT CHECK-IN
------------------------------------	------------------------------------	--------------------------------------

Ready for a transformation like Corvale Pharmaceuticals's?

Emerson Technology designs and builds the data, AI, cloud and automation systems behind results like these. On a free consultation call we will map this exact playbook to your business - your data, your constraints, your numbers.

Book a free consultation

support@emersontechnology.com

or start online: emersontechnology.com/contact