LIFE SCIENCES · DATA GOVERNANCE

A Global Biotech Enterprise

Stopping the bleeding on an unknown data liability.

How a global biotech turned a DOJ Executive Order 14117 compliance obligation into a continuously operated Trust Layer, proven inside the company’s own production environment before the program scaled across the enterprise.

FOREWORD

The situation, in short.

A global biotech engaged Data Sentinel in 2025 with a regulatory clock already running and no clear picture of what was inside its own environment. Its own statement of work described the compliance liability as of “unknown volume and severity.”

What follows is how the Trust Layer was proven inside the client’s estate before a multi-year commitment was made. Not a discovery report. Not a remediation roadmap. A platform running inside the environment, reducing risk as it scaled.

“We didn’t want another report telling us how exposed we were. We wanted to see our Covered Data, govern it automatically, and know it would keep pace as the rules change.”

DIRECTOR OF DATA AND ANALYTICS GOVERNANCE

CHAPTER ONE

An unknown liability, with a deadline.

Executive Order 14117 was signed in February 2024. The Department of Justice published the Final Rule implementing it in January 2025. Compliance phased in through the year.

The rule restricts covered transactions that give countries of concern access to bulk sensitive personal data on U.S. persons. Six countries are named. China is one of them.

Covered Data spans several categories: personal identifiers, precise geolocation, biometric identifiers, human genomic data and biospecimens, other human ‘omic data, personal health data, and personal financial data. Each has its own threshold defining what counts as “bulk.” Cross a threshold in a covered transaction and the transaction is either restricted or prohibited outright.

> **THE THRESHOLD THAT MATTERS**
>
> Human genomic data crosses into bulk at just 100 U.S. persons. That one number is why life sciences sits at the sharp end of this rule.

A global biotech running commercial and research across both the United States and China is squarely in scope. Genomic and clinical data on U.S. persons is exactly the kind of Covered Data DOJ singled out. Cross-border data movement is exactly the kind of transaction the rule constrains.

When the client engaged Data Sentinel, the compliance liability was, in the client’s own words, of **“unknown volume and severity.”** They did not know how much Covered Data they held. They did not know where it sat. They did not know which cross-border flows carried it. The regulatory clock had already started.

CHAPTER TWO

Why the discovery report model doesn’t reduce risk.

The default playbook for a compliance obligation of this shape is a discovery project. A consulting firm scopes the assessment, interviews stakeholders, samples systems, produces a classification report, delivers a remediation roadmap, and closes the engagement. Remediation gets scoped later, budgeted later, staffed later.

That model has two problems, and both hit exactly where this regulation lives.

The first is that the report is out of date the day it ships. Data keeps entering the environment after the sample window closes. Retention decisions get made or not. Regulatory definitions evolve. The classification snapshot describes an environment that no longer exists.

The second is worse. A classification report tells the company how much Covered Data was there. It does not reduce how much is there today. The exposure sits in a binder waiting for a project to be funded and staffed. Every day of that gap is a day the enforcement risk continues at its measured level, not below it.

Data Sentinel does not run discovery projects. The platform is deployed inside the client’s own environment on day one and starts classifying immediately. Classification is continuous, not a snapshot. Evidence flows out as a byproduct of the platform running, not as a separate audit-prep exercise. Risk reduction starts the moment the classifier is on.

That is what changed at the client.

CHAPTER THREE

Prove it live, then scale.

The client did not want a multi-year commitment before they saw the platform work inside their own environment. That constraint shaped how the engagement was designed.

Data Sentinel proposed a proof-gated deployment. The platform would run inside the client's actual production estate, not a sandbox. It would run against three sources chosen to span the environment's diversity: one structured enterprise data warehouse, one unstructured collaboration platform, and one analytics lakehouse. Each source type presents a different classification challenge. Getting all three right meant the platform could handle the rest.

Before any scan touched production data, the label taxonomy had to be aligned with the client's data classification governance and privacy office. Data-owner sign-off was required. Automated classification at scale, without governed labels, would have created expensive reclassification rework months later. Governance sequencing wasn't a design preference. It was the only way the automation would hold up.

The proof-gate had specific success criteria. Classification accuracy across the four DOJ Covered Data categories had to clear a threshold. The platform had to run inside production without disrupting normal operations. Data-owner sign-off had to be maintained through the scan cycle. Three sources. Three success criteria per source.

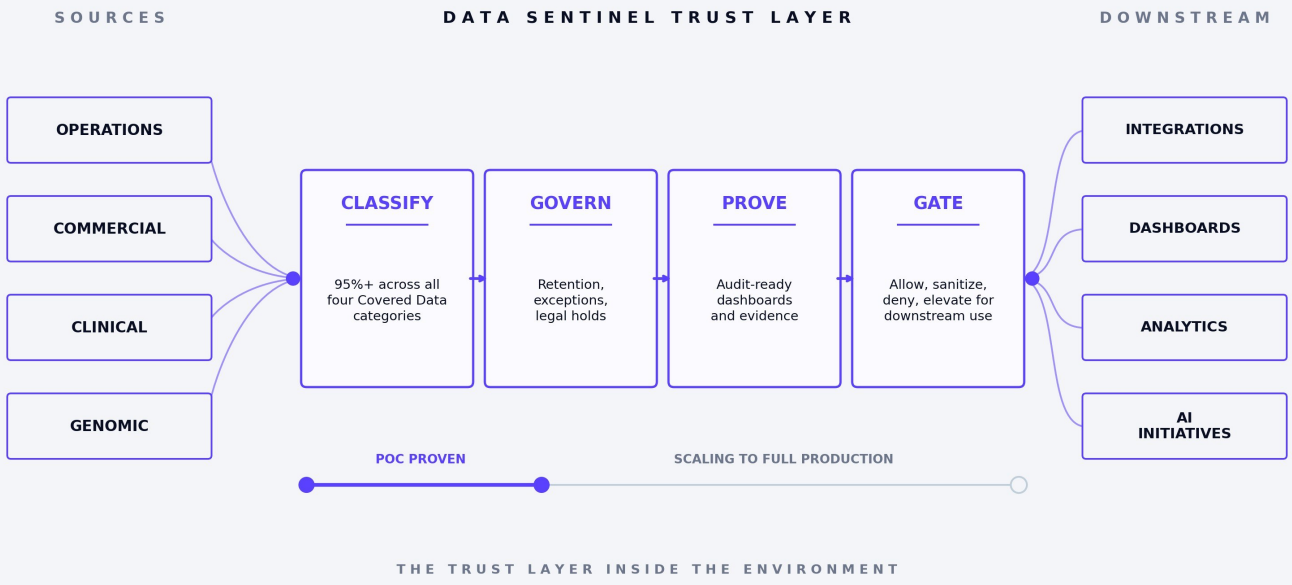
Every one cleared.

The proof also carried a second effect. While the POC ran, aggregate risk on net-new Covered Data started dropping, because the classifier was already flagging what was in-scope as it arrived. The reduction was real, not projected. That reality reframed the client's decision. Scaling was no longer a bet on whether the platform would work. It was extending something that already was.

CHAPTER FOUR

The Trust Layer inside the environment.

The Trust Layer at the client runs as a pipeline inside their environment. Data enters unclassified, passes through four stages, and exits ready for downstream use. Evidence is produced at every step.



The Trust Layer pipeline. Data enters from operational systems, passes through Classify, Govern, Prove, and Gate stages, and exits ready for downstream consumption.

Classify. The self-learning engine runs at 95%+ accuracy across all four DOJ Covered Data categories. Personal identifiers. Genomic and biospecimen data. Biometric and health data. Financial and location data. Each labeled as it moves. Data-owner-reviewed exceptions get flagged.

Govern. Retention, lifecycle, and exception handling. Policies aligned to the client's classification governance are enforced as data moves through the environment. Legal holds get labeled and preserved. Retention timers run against what is stored.

Prove. Audit-ready evidence produced continuously. Dashboards report on what has been classified, what has been enforced, what has been remediated. When the auditor asks, the evidence is already there.

Gate. Every downstream request evaluated. Allow. Sanitize. Deny. Elevate for review. Analytics, dashboards, integrations, and AI initiatives only see what has cleared the pipeline.

CHAPTER FIVE

What the POC proved.

The POC cleared every success criterion across every source type. The specific results:

- **95%+ classification accuracy** across all four DOJ Covered Data categories.
- Three representative production sources cleared under proof gating.
- Zero disruption to client operations during scan cycles.
- Data-owner sign-off maintained throughout the engagement.
- Aggregate risk on net-new Covered Data dropping continuously as the classifier ran.

What that opens up now that the platform is scaling beyond proof:

- A **governed AI data pipeline.** Analytics and AI running against classified, policy-eligible data with full auditability.
- A **continuously shrinking DOJ exposure.** Net-new risk is already lower. Historical exposure drops next.
- A **posture that absorbs regulatory change** without re-procurement. As DOJ definitions evolve, the classifier retunes.
- Cross-border U.S. and China research that **keeps moving while staying compliant.**

CHAPTER SIX

From triage to managed service.

The engagement is structured as three sequential phases. Each builds on what has already been proven.

Phase One. Triage.

Full production rollout beyond the three POC sources. Identity and single sign-on integration. Expanded connectivity to the next wave of sources across clinical trial systems and clinical operations platforms. ITSM workflow integration goes live so classification exceptions and remediation actions flow through the client’s existing operational rhythm. Governance activated across the environment. This is where the POC scales from three sources to the full estate.

Phase Two. Assessment and Remediation.

The bottoms-up historical scan begins. Everything the platform did not touch in Phase One gets classified. Historical Covered Data that has been accumulating for years finally gets sized. Remediation gets prioritized. The historical iceberg gets measured, then reduced. The Trust Layer that stopped the bleeding is now the mechanism for shrinking the accumulated liability.

Ongoing. Managed Service.

Long-term managed-service operation with monthly strategic reviews. At that point, DOJ 14117 compliance is not a project the client runs. It is a state the environment maintains.

Prove it. Scale it. Manage it. Reduce risk at every step.

ABOUT DATA SENTINEL

Trusted data. On day one. Every day after.

Data Sentinel builds continuous Trust Layers for regulated enterprises. Our platform sits inside customer environments, classifying, governing, and producing evidence continuously.

Compliance, privacy, and AI readiness become byproducts of daily operation, not separate projects. When regulations evolve, our platform adapts. When new data enters the environment, it is classified as it lands. When auditors ask, the evidence is already there. That is what a Trust Layer does that a report cannot.

Protect

Sensitive data, continuously classified.

Comply

Evidence produced as operation runs.

Govern

Policies enforced at scale, live.

Grow

AI launched against trusted data.

HOW TRUSTWORTHY IS YOUR DATA?

Book a demo to see how a **continuously operated Trust Layer** *becomes the foundation for compliance, quality, and AI.*

data-sentinel.com

Data Sentinel builds continuous Trust Layers for regulated enterprises across life sciences, aviation, insurance, energy, and financial services.