NonStop

Clinical Lab Data Maturity Assessment

Is Your Lab's Data Infrastructure Ready to Scale?

If you're a Lab Director, CTO, or Head of Bioinformatics, you've felt this pain:

- Your genomic, clinical, and operational data is scattered across disconnected systems — LIMS, EHR, sequencers, research databases — with no single source of truth.
- Your scientists spend 40%+ of their time wrangling data instead of analyzing it.
- Data pipeline failures go undetected until dashboards break or reports are wrong.
- You can't answer basic questions like "What's our true cost-per-sample?" or "Are we audit-ready?" without weeks of manual work.
- AI/ML initiatives stall because your proprietary genomic data is trapped in silos, unlabeled, and unstructured.

Take our 5-minute Clinical Lab Data Maturity Assessment and discover exactly where your data architecture stands — and what to fix first.

The Assessment Questions

Section 1: Data Integration (4 questions)	Score
Q1: How many separate systems house your lab's critical data today?	
☐ 1-3 systems	25
☐ 4-5 systems	15
☐ 6-8 systems	08
☐ 9+ systems	0
Q2: How is data currently shared between these systems?	
☐ **Real-time, automated data sync** via APIs or event-driven integration (HL7/FHIR, Kafka, etc.) — systems talk to each other without manual intervention.	25
☐ **Scheduled batch jobs** — ETL pipelines run nightly or weekly to sync data across systems, but there's lag.	15
☐ **Manual exports and imports** — team members export CSVs or Excel files from one system and upload to another.	08
☐ **No integration** — data stays siloed in each system; reconciliation happens ad hoc when needed (or not at all).	0
Q3: Do you have a single source of truth for core entities (Patients, Providers, Orders, Tests)?	
☐ **Yes** — We have one authoritative system or unified data layer (e.g., a data warehouse or master data hub) where Patient, Provider, Order, and Test records are reconciled and governed.	25
☐ **Partially** — We have a single source of truth for some entities (e.g., Patients), but not all. Or our "source of truth" is only as current as the last batch sync.	12
☐ **No** — Each system maintains its own version of Patients, Providers, Orders, and Tests. There's no unified master record.	0
Q4: When was the last time a data discrepancy between systems caused a real business problem?	
☐ **Never / Can't recall** — Our data is consistent across systems, and discrepancies are caught before they cause issues.	25
☐ **More than 6 months ago** — It's happened, but it's rare.	15
☐ **Within the last 1-6 months** — We've had at least one incident where mismatched data (duplicate records, wrong order status, incorrect patient info) caused downstream problems.	05
☐ **Within the last 30 days** — This is a recurring problem. Data discrepancies between systems cause operational headaches, report errors, or compliance risk regularly.	0

Section 2: Pipeline Reliability (3 questions)	Score
Q5: How often do your data pipelines (ETL, bioinformatics workflows, report generation, etc.) fail or require manual intervention?	
☐ **Rarely / Never** — Pipelines run reliably without manual babysitting. Failures are rare (< 1% of runs) and automatically retried.	25
☐ **Occasionally** — Pipelines fail a few times per month, but we catch and fix them quickly.	15
☐ **Weekly** — We're troubleshooting or manually re-running pipelines at least once a week.	05
☐ **Daily / Multiple times per week** — Pipeline failures are a daily occurrence. Our engineering team spends significant time firefighting instead of building.	0
Q6: If your sample/test volume doubled tomorrow, what would happen to your data pipelines?	
☐ **They'd handle it seamlessly** — Our pipelines are designed to auto-scale (cloud-native, elastic compute). We've tested this.	25
☐ **They'd handle it with minor adjustments** — We'd need to bump up resources or tweak some configs, but no major rework required.	15
☐ **They'd struggle** — We'd see slower processing, longer TAT, and likely some failures. We'd need to do significant re-architecture.	05
☐ **They'd break** — Our current pipelines can't handle the volume we have today reliably, let alone 2x.	0
Q7: Do your bioinformatics or data pipelines have automated quality checks (QC gates) at each stage?	
☐ **Yes, fully automated** — Every stage of our pipelines has automated QC checks, and bad data is flagged or rejected before it reaches clinical reports or dashboards.	25
☐ **Partially automated** — Some stages have automated QC in pipelines, but manual review happens before reports go out.	12
☐ **Mostly manual** — QC happens, but it's done by humans reviewing outputs after the fact.	05
☐ **No systematic QC** — We don't have formal QC gates built into our pipelines.	0

Section 3: Real-Time Data Access (2 questions)	Score
Q8: How long does it take for a critical lab event (e.g., sample QC failure, urgent result, pipeline completion) to show up in the systems or dashboards that need it?	
☐ **Seconds to minutes (real-time)** — Events are pushed immediately via streaming, webhooks, or event-driven architecture.	25
☐ **Within an hour** — Near-real-time via frequent batch syncs (every 15-60 min).	15
☐ **Several hours** — Batch jobs run every 4-12 hours, so there's noticeable lag.	05
☐ **Next day or longer** — Data syncs happen nightly or less frequently. Decisions are made on stale data.	0
Q9: Can your lab operations team see real-time status of samples, pipelines, and test orders — and act on it without delays?	
☐ **Yes** — We have live dashboards showing sample location, pipeline run status, QC pass/fail, and order fulfillment in real time. Data is available when decisions need to be made.	25
☐ **Partially** — We have dashboards, but they're updated in batches (hourly or daily). Occasional decision delays happen due to data lag.	12
☐ **No** — Our team has to check multiple systems or run manual queries to get status updates. Decision delays due to missing data are frequent.	0

Section 4: Governance & Compliance (4 questions)	Score
Q10: If a regulator (CAP, CLIA, FDA) or auditor asked you to show full data lineage for a specific patient result — from sample receipt to final report — how quickly could you produce it?	
☐ **Within minutes** — We have automated lineage tracking. We can pull a full audit trail (who touched the data, when, what changed) instantly.	25
☐ **Within a few hours to a day** — We can reconstruct it by querying logs and systems, but it requires manual effort.	12
☐ **Several days** — We'd need to dig through multiple systems, reconcile records, and piece together the story.	05
☐ **We can't reliably produce full lineage** — Our current systems don't capture all the steps or changes.	0
Q11: How is PHI (Protected Health Information) / sensitive genomic data currently protected in your non-production environments (dev, QA, analytics)?	
☐ **Automated controls** — PHI is automatically masked, or we use synthetic/de-identified datasets. RBAC is enforced at the data layer.	25
☐ **Manual processes** — We manually create sanitized exports or de-identified copies for non-production use.	12
☐ **Production data in dev/QA or no formal controls** — Developers and analysts work directly on real PHI, or PHI handling isn't strictly governed.	0
Q12: When was your last compliance audit (CAP, CLIA, HIPAA, SOC 2), and how did it go?	
☐ **Within the last 12 months, and we passed with no major findings** — We're audit-ready and maintain continuous compliance.	25
☐ **Within the last 12 months, with minor findings** — We passed, but had to remediate a few gaps.	15
☐ **More than a year ago, or we had significant findings** — The audit revealed compliance gaps that required substantial rework.	05
☐ **We haven't been audited yet, or we failed** — We're not confident we'd pass an audit today.	0
Q13: Do you have a formal data governance framework in place (data ownership, stewardship, access policies, metadata management)?	
☐ **Yes, fully implemented** — We have documented policies, assigned data owners/stewards, a metadata catalog, and enforced access controls.	25
☐ **Partially implemented** — We have some policies and tools, but governance isn't consistently enforced across all datasets.	12
☐ **In planning / pilot phase** — We know we need it, and we're starting to build it.	05
☐ **No formal framework** — Data governance is ad hoc. Policies aren't documented or enforced.	0

Section 5: AI/ML Readiness (2 questions)	Score
Q14: Can your data science/bioinformatics team access a unified, analysis-ready dataset of your proprietary genomic, clinical, and phenotypic data today?	
☐ **Yes** — We have a governed data warehouse or lakehouse where all our omic + clinical data is unified, labeled, and ready for analysis or model training.	25
☐ **Partially** — Some datasets are unified, but scientists still have to join or clean data from multiple sources for most projects.	12
☐ **No** — Our data is scattered. Every AI/ML project starts with months of data wrangling and ETL work.	0
Q15: How much time does your team spend on data preparation vs. analysis in a typical AI/ML project?	
☐ **<20% on data prep** — Our data is already clean, labeled, and structured. Teams spend most of their time on modeling and insights. Score: 25	25
☐ **20-40% on data prep** — Data prep is part of the workflow, but it's manageable. Score: 15	15
☐ **40-60% on data prep** — Data wrangling is a major bottleneck. Score: 5	05
☐ **>60% on data prep** — Scientists spend the majority of their time just getting data into a usable state. Score: 0	0

Maturity Tiers

Based on your overall score, you'll be classified into one of four maturity tiers:

Score Range	Maturity Tier	What This Means
0-40	Foundational	Critical infrastructure gaps holding you back from scaling and compliance
41-65	Emerging	Solid progress made, but key weaknesses preventing you from reaching best-in-class
66-85	Advanced	Strong infrastructure with optimization opportunities in real-time access and AI deployment
86-100	Leading	World-class data infrastructure; ready to monetize and lead the industry