The Foundation of Trust:

Pathology Governance and Compliance in the Age of AI

A joint perspective from Proscia and PIROuette Dx on building confidence in digital pathology through responsible data practices and transparent AI

PROSCIA®

PIROuette Dx™

Turning Insight into Truth

* **BY: HANIYYAH BONVINI, PHARMD, BCPS**
DIRECTOR, DATA GOVERNANCE & COMPLIANCE
PIROUETTE DX

Pathology is undergoing a profound transformation. The shift from glass slides to digital workflows has unlocked new opportunities for collaboration, quality improvement, research, and AI development. Whole slide images (WSIs), metadata, annotations, and clinical records now form one of healthcare's richest data ecosystems. As this ecosystem grows, so do expectations around both its advantages and its stewardship.

For decades, data governance in healthcare was treated as a necessary burden — a compliance-oriented function designed to satisfy regulators, minimize liability, and prevent what could go wrong. Governance lived primarily within IT and compliance teams, and success was measured by audits passed and penalties avoided.

Now, as pathology digitalizes, governance has become a strategic capability rather than an administrative constraint. WSIs, metadata, annotations, and derived datasets are no longer static; they are reusable, high-value assets that power clinical operations, research, real-world evidence, and AI development.

For CIOs and institutional leaders, this shift is fundamental. Governance now determines how quickly data can be mobilized across sites, how safely it can be reused for research and AI development, how confidently organizations can engage external partners, and how effectively risk can be managed without slowing innovation. Poor governance not only creates compliance exposure, but also stalls strategic initiatives.

As one of the most data-rich and diagnostically critical domains in healthcare, pathology sits at the heart of this transformation. Departments now manage petabyte-scale image archives distributed across on-premise, cloud, and hybrid environments — each decision introducing new implications for security, interoperability, and long-term value creation.

This whitepaper outlines a modern framework for pathology data governance, compliance, and operational readiness. It is designed for pathology leaders, medical directors, CMOs, CIOs, and institutional decision-makers who must navigate a complex regulatory landscape while preparing their organizations for the next era of diagnostic innovation.

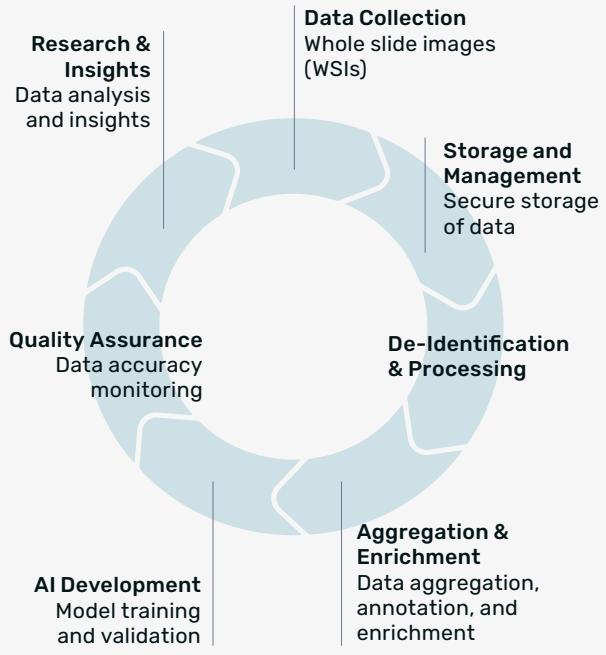
This paper provides leaders with:

- A clear definition of how pathology data functions today — operationally and as a strategic asset
- A governance model used by leading organizations, supported by roles, committees, standards, and cross-functional alignment
- An overview of the legal and regulatory landscape that shapes responsible digital pathology
- Operational best practices that turn governance into daily reality
- A roadmap for preparing high-integrity data environments that support responsible AI adoption

Digital pathology's future will be shaped not by the speed of technology deployment, but by the strength of the data foundations beneath it. The institutions that embrace governance today will be the ones best prepared to adopt AI safely, ethically, and effectively tomorrow.

> "Every stage of the pathology data lifecycle is an opportunity to strengthen — or weaken — trust."
>
> **GIOVANNI LUJAN, MD, FCAP**
> PROFESSOR AND VICE CHAIR, PATHOLOGY INFORMATICS AND COMPUTATIONAL PATHOLOGY, THE OHIO STATE UNIVERSITY

FIGURE 1
DATA LIFECYCLE IN PATHOLOGY



Pathology has always been central to diagnosis. What has changed is the value pathology can have beyond 1:1 patient diagnosis. As organizations transition to digital slide management, integrated viewers, analytics platforms, and increasingly AI-augmented workflows, pathology is becoming a data-intensive discipline. Every case now generates digital artifacts that extend far beyond the traditional glass-slide workflow.

Digital pathology introduces new scale, complexity, and responsibility—making governance, compliance, and operational rigor essential. Departments now manage terabytes to petabytes of sensitive data, often distributed across systems, vendors, and teams. What previously lived entirely under a microscope and in storage cabinets now lives in cloud environments, and research pipelines — and each step introduces new opportunities and responsibilities.

The path to that future begins with clear definitions of data use, governance, and compliance — and a shared understanding of the environment pathology leaders now operate within.

"We're at an inflection point. For the first time in a century, the limiting factor in pathology isn't what we can see. It's how well we manage what we've digitized."

MATTHEW CECCHINI, MD, PHD
ASSOCIATE PROFESSOR,
WESTERN UNIVERSITY, LONDON ONTARIO CANADA

03. DEFINING PATHOLOGY DATA USE, GOVERNANCE & COMPLIANCE

Digital pathology has expanded both the volume of data labs generate and the stakes of using it. To manage this ecosystem responsibly, teams need a shared understanding of three core concepts: data use, data governance, and compliance. These terms form the backbone of safe clinical operations, ethical research, and AI readiness.

03.01 How Pathology Data Is Used

Pathology data now serves multiple purposes, each with its own responsibilities and risks

Operational Use supports primary diagnosis, peer review, quality assurance, and multidisciplinary discussions. These uses require data that is complete, accurate, timely, and reliably accessible within secure clinical systems.

Secondary Use has grown quickly as organizations increasingly reuse diagnostic data for research, education, workflow analysis, real-world evidence generation, and early-stage AI development. These activities introduce different ethical and regulatory considerations, particularly when real patient data is repurposed outside direct clinical care.

AI-Driven Use is an emerging category where images, metadata, and annotations become inputs to model training, validation, and monitoring. These workflows require more than availability; they demand well-curated data with clear provenance, consistent metadata, and rigorous documentation.

As captured in the data lifecycle graphic (Figure 1), these categories often overlap — and without clear policies, the very same dataset can be safe in one context and risky in another. That is what makes governance essential.

03.02 What Governance Really Means

If data use describes what pathology teams do with their data, governance describes how that data is stewarded across its entire lifecycle. Governance ensures that data remains trustworthy whether it is used for diagnosis, shared for quality initiatives, or prepared for research and AI.

Governance has evolved from a compliance function into a strategic capability that enables safe reuse across clinical care, research, and AI. It is no longer just about preventing what shouldn't happen—it is about enabling what must happen next.

At its core, governance exists to maintain data integrity, reduce risk, and create clarity across teams. In well-governed organizations, responsibilities are shared rather than siloed. Clinical leaders retain ownership of diagnostic correctness and appropriate use; informatics and operations teams act as stewards of metadata quality and documentation; and IT leaders serve as custodians who maintain secure systems, enforce access controls, and protect the infrastructure.

"Disciplined governance eliminates confusion. When roles are precise and decisions are guided by clear policy, momentum increases."

M. E. DE BACA, MD
VICE PRESIDENT FOR MEDICAL AFFAIRS, SYSMEX AMERICA, GOVERNOR, COLLEGE OF AMERICAN PATHOLOGISTS (CAP), CHAIR, CAP COUNCIL ON INFORMATICS AND PATHOLOGY INNOVATION

Leading health systems increasingly recognize clinical data as a non-depleting institutional asset. When governed effectively, high-quality, longitudinal pathology data fuels a virtuous cycle: better data enables better analytics and AI; better insights improve clinical care and operations; and improved care generates even higher-quality data. At the same time, external demand from pharmaceutical companies, AI developers, and research collaborators has elevated governance to a board-level concern—one tied directly to research competitiveness, revenue diversification, and long-term relevance.

Mature governance frameworks to manage this asset are grounded in principles such as accountability, transparency, interoperability, standardization, auditability, and privacy-by-design. These principles allow organizations to protect patient trust while still unlocking value at scale. Governance transforms data from isolated artifacts locked inside departments into a shared, reusable, and AI-ready institutional resource.

The Governance Maturity Model (Table 1) illustrates this evolution—from ad hoc, compliance-driven oversight to proactive, enterprise-wide governance that supports analytics, interoperability, and AI development. Organizations that treat governance as compliance-only often struggle to scale reuse and AI safely—not because they lack effort, but because the model wasn't built for modern data volume and complexity.

OPERATIONALIZATION CALLOUT (PIROUETTE DX)
GOVERNANCE STARTER PACK

If you want governance to work in a real pathology workflow, assign these six decisions explicitly and document them in a one-page charter:

1. Who is the Clinical Data Owner for secondary use (final clinical accountability)?
2. Who is the Data Steward (owns schemas, templates, logs, and day-to-day traceability)?
3. Who is the Data Custodian (access, security controls, audit logs, retention)?[1] [2]
4. Who can authorize release (QA sign-off, PHI checks, and packing list reconciliation)?[3] [4]
5. Who can approve re-pulls/longitudinal linkage (and under what IRB scope)?
6. Who leads incidents and Corrective and Preventive Actions (CAPA) closure (quality/compliance authority)?

Defining roles and answering these questions up front prevent downstream delays and ad hoc exceptions.

1 College of American Pathologists. Retention of laboratory records and materials. Accessed March 6, 2026. https://www.cap.org/member-resources/practice-management/strategy-and-planning

2 DICOM Standards Committee. DICOM whole slide imaging (WSI). Accessed March 6, 2026. https://dicom.nema.org/dicom/dicomwsi/

3 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/

4 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html subchapter-C/part-164/subpart-E/section-164.514

TABLE 1
GOVERNANCE MATURITY MODEL

	SUBOPTIMAL	MANAGED	OPTIMIZED	AI-READY
DESCRIPTION	Governance practices are ad hoc and inconsistent. Data is often siloed with limited oversight.	Foundational governance policies and accountability structures are established.	Governance is standardized and proactive across the organization.	Governance supports advanced analytics and AI development with ethically managed, interoperable data.
KEY CHARACTERISTICS	• Minimal documentation • Reactive compliance checks • Department-level ownership	• Documented policies • Defined data stewards • Periodic audits and access reviews	• Cross-functional governance committee • Automated data validation • Regular quality audits	• FAIR data principles (Findable, Accessible, Interoperable, Reusable) • Bias monitoring and explainability frameworks • Continuous oversight of AI models
EXAMPLE INDICATORS	• Dispersed data storage • Limited audit trails	• Governance charter in place • Routine compliance training	• Central data catalog • Audit-ready metadata	• Federated data architectures • AI governance incorporated into SOPs

"Compliance is the floor; governance is the framework."

GUILLERMO G. MARTINEZ-TORRES, MD, FCAP
PRESIDENT & CHIEF PHYSICIAN EXECUTIVE
NORDX LABORATORIES, MAINEHEALTH

03.03 Compliance: Federal, State, International, and Organizational Guardrails That Keep Data Safe

Compliance complements governance by defining the legal and ethical boundaries within which pathology data must be handled. In the United States, HIPAA and the HITECH Act govern privacy, security, and breach requirements, while CLIA and the College of American Pathologists (CAP) define expectations for diagnostic quality and digital record retention. The FDA overlays additional guidance for digital pathology systems, clinical decision support tools, and emerging AI technologies. These are all discussed in more detail in section 06.01, Federal Requirements: Protecting Privacy and Diagnostic Integrity.

Federal privacy protections and requirements are increasingly supplemented by state laws, each with its own set of requirements. States such as California (CCPA/CPRA), Colorado, Connecticut, Utah, and Virginia have introduced privacy laws that impose specific expectations around consent, data minimization, patient rights, and de-identification. State-specific laws often extend beyond HIPAA, affecting how pathology data can be used for research, training, or analytics—particularly when secondary uses involve commercial collaborators or AI development. For multi-state institutions, this creates a patchwork of obligations that must be harmonized through strong internal governance. Governance ensures that policies remain uniform even when regulations differ.

Within institutions, research oversight — including IRBs, data use agreements, and federally defined de-identification standards — provides additional control over how diagnostic data may be reused for secondary purposes. These frameworks ensure transparency and protect patients when data moves beyond clinical care.

Finally, international regulations such as General Data Protection Regulation (GDPR) and the EU AI Act increasingly influence collaborative research and AI development. Even U.S.-based organizations feel the impact as vendors and partners adopt global best practices to meet these requirements.

TABLE 2
RELEVANT REGULATIONS AND IMPACT

AREA	EXAMPLES	KEY FOCUS	IMPACT ON DIGITAL PATHOLOGY
FEDERAL HEALTHCARE REGULATIONS (U.S)*	HIPAA, HITECH	Privacy, security, breach notification[5]	Governs how patient data is stored, accessed, and protected
LABORATORY & DIAGNOSTIC STANDARDS*	CLIA, CAP	Diagnostic quality, record retention	Sets expectations for digital workflows and long-term data management
FDA OVERSIGHT*	FDA guidance	Digital pathology systems, clinical decision support (CDS) tools, AI	Influences validation, deployment, and use of AI-driven technologies
STATE PRIVACY LAWS	CA (ex: CCPA / CPRA), CO, CT, VA	Consent, minimization, patient rights, de-identification	Creates varying requirements across states that must be harmonized
INSTITUTIONAL RESEARCH OVERSIGHT	IRBs, data use agreements (DUAs), de-identification standards	Secondary data use, transparency	Additional control over how clinical data is reused for research and innovation
INTERNATIONAL REGULATIONS	GDPR, EU AI Act	Data protection, AI governance	Impacts global collaboration and vendor best practices

* Discussed in further depth in Section 06.01.
Note: This table highlights key regulatory areas but is not an exhaustive list of all applicable regulations.

03.04 Putting Definitions Into Action

Data use, governance, and compliance are decisions that need to be made and operationalized early. The table below outlines what leading Academic Medical Centers (AMCs) typically accomplish in their first 60–90 days to translate these definitions into a defensible foundation for digital pathology.

TABLE 3
A STRONG FOUNDATION IN 60-90 DAYS

IN THE FIRST 60-90 DAYS, LEADING AMCS TYPICALLY:	
ASSIGN DECISION RIGHTS	Confirm data owners, stewards, and custodians across pathology, operations/informatics, and IT/security.
MAP THE DATA LIFECYCLE	Document where WSIs and metadata are created, stored, accessed, and archived (clinical and secondary use).
STAND UP A REQUEST WORKFLOW	Implement an intake + approval path for research / secondary use with consistent documentation and auditability.
DEFINE MINIMUM STANDARDS	Choose a target for WSI format (e.g., DICOM-WSI) and a minimum metadata set required for reuse.
OPERATIONALIZE SECURITY BASICS	Enforce role-based access, encryption at rest/in transit, and logging for key systems and repositories.
PICK 2-3 PRIORITY USE CASES	Focus on one clinical workflow (e.g., remote consults), one QA workflow, and one research dataset request as “lighthouse” wins.
SET GOVERNANCE CADENCE	Schedule a recurring review (monthly is enough early on) to resolve exceptions and refine policies as adoption scales.
OUTCOME	A DEFENSIBLE FOUNDATION FOR DIGITAL PATHOLOGY THAT REDUCES RISK AND ACCELERATES TIME-TO-VALUE.

5 US Department of Health and Human Services. Breach Notification Rule. https://www.hhs.gov/hipaa/for-professionals/breach-notification/index.html

The modern pathology environment is now a distributed data ecosystem spanning acquisition, storage, access, and analytics. This section breaks down the technical foundation beneath modern digital pathology: where data lives, how it's protected, and what standards are required to keep workflows interoperable and AI-ready.

"Technology alone doesn't make you ready. AI-ready pathology is built on intentional governance, operational discipline, and cultural trust."

MARILYN BUI, MD, PHD
SENIOR MEMBER & PROFESSOR OF PATHOLOGY,
MOFFITT CANCER CENTER

04.01 Infrastructure Models

Pathology leaders today face decisions about where their data should live and how it should flow. Although infrastructure decisions often begin as IT questions, they quickly become operational, compliance, and strategic questions, as well. There are three primary infrastructure models: on-premises, cloud-based, and hybrid.

Infrastructure models are particularly important to consider as the volume of pathology data continues to grow: whole-slide images (WSIs) often range from 2-4 GB each. At 700–1,000 slides per day, that translates to roughly 3 terabytes per day and nearly one petabyte per year for a mid-large pathology program.[6]

On-Premises Infrastructure

On-premises infrastructure offers local control — servers sit within the organization's data center, network access can be tightly managed, and IT teams are familiar with support patterns. But that control comes with trade-offs: significant capital investment, ongoing maintenance, hardware refresh cycles, and limited scalability as image volumes grow. For institutions with multiple sites, remote access can strain bandwidth and degrade reading performance. On-prem remains viable for small programs or early-stage digital adoption, but its limitations become pronounced as organizations scale or prepare for AI.

Cloud-Based Infrastructure

Cloud adoption has accelerated in digital pathology due to its flexibility, scalability, and resilience. In a cloud-native environment:

- Storage expands elastically
- Compute resources scale with user demand
- Redundancy and disaster recovery are built-in
- Remote access becomes seamless
- Zero-trust and modern security frameworks come standard

Cloud platforms can offer advanced security controls and resilience that many organizations would otherwise need significant investment to match, especially when designed with healthcare-grade governance. Table 4 reinforces that cloud environments are increasingly the backbone of AI-ready pathology programs.

"The cloud doesn't just store pathology data — it unlocks it. With the right foundation, you can scale access, protect patient trust, and turn everyday cases into long-term clinical and research value."

GIOVANNI LUJAN, MD, FCAP
PROFESSOR AND VICE CHAIR, PATHOLOGY INFORMATICS AND
COMPUTATIONAL PATHOLOGY, THE OHIO STATE UNIVERSITY

6 Lee J, Takemaru L, Bappy DM, et al. Adaptive compression framework for giga-pixel whole slide images. Nat Commun. 2025. https://www.nature.com/articles/s41467-025-66889-0

Hybrid Infrastructure

Some organizations find themselves in some stage of hybrid architecture — combining local file systems, cloud-based archives, and multiple clinical systems. Hybrid models often emerge organically as organizations modernize, migrate workloads, or onboard new tools.

However, hybrid environments can become complex if not managed intentionally: data may flow between systems on different standards, duplicated datasets may proliferate, and workflows may rely on brittle integrations. Hybrid is often a transitional step, not an end state — and to succeed with it, organizations need clear governance, architectural discipline, and a roadmap for future consolidation.

TABLE 4
DIGITAL PATHOLOGY INFRASTRUCTURE

	ON-PREMISE	HYBRID	CLOUD
IT RESOURCES REQUIRED	High internal staff & maintenance	Moderate; shared responsibilities	Minimal internal IT
UPFRONT COSTS	Capital expenditure (hardware, servers)	Medium	Very low
SCALABILITY	Limited, slow to expand	Flexible based on mix	Near-infinite, fast
SECURITY & COMPLIANCE	Full control locally	Shared control models	Leading cloud certifications & controls
IDEAL FOR…	Labs with strict data residency & bespoke workflows	Labs modernizing gradually or with mixed environments	Labs needing agility, scalability, reduced overhead

04.02 Safeguarding the Digital Pathology Ecosystem

As data volumes scale and infrastructures become more distributed, security becomes both more important and more complex. "Zero-trust" security models are the standard: pathology environments must assume no implicit trust between systems, users, or connections.

Effective safeguards include:

- Multi-layer encryption (at rest and in transit)
- Role-based and attribute-based access control
- Continuous auditing and activity logging
- Network segmentation
- Regular penetration testing
- Disaster recovery and multi-region backups

Security is not solely an IT concern. It directly affects diagnostic quality, patient privacy, and institutional credibility. Because WSIs are enduring clinical records, they require durable safeguards—strong access controls, encryption, and auditing that protect patient trust over time.

04.03 Data Standards: Building for Interoperability & AI

The ability to move pathology data cleanly between systems, sites, and teams is the foundation of digital operations, research collaboration, and future AI development. Yet many pathology departments still rely on proprietary formats, inconsistent metadata, or local conventions that were never designed for large-scale digital workflows.

When a case moves from one system to another — from scanner to viewer, from viewer to LIS, from LIS to research environment — each system makes assumptions about the data: how it is labeled, how it is structured, what metadata fields exist, how magnification is stored, or whether annotations are attached. With standards, every handoff becomes seamless — reducing friction, simplifying operations, and directly improving readiness for research and AI. For leaders, the impact becomes clear:

- Less time troubleshooting and reformatting data
- More reliable multi-site operations
- Greater confidence in research datasets
- Easier onboarding of new tools or partners
- A direct path toward AI readiness

DICOM-WSI: A Standard for Images

DICOM – the Digital Imaging and Communications in Medicine standard – provides the standard for whole slide imaging.

Governance mandates are shifting away from vendor-specific formats (e.g., .svs, .ndpi) and toward DICOM formats to ensure long-term readability. This has been used in radiology for decades, and its adoption in digital pathology means that:

- Images and metadata travel together, so context is never lost
- Vendors become interoperable, reducing long-term dependency
- AI developers get consistent, machine-readable structure
- Health systems can centralize storage instead of juggling formats, and create Enterprise Imaging Governance that includes radiology images alongside WSIs

The point for clinical and operational leaders is not the file format itself. It's that a scan created today will still be usable, searchable, and trustworthy ten years from now, no matter how systems evolve.

Studies published in the Journal of Pathology Informatics and DPA/FDA workshops show that a majority of institutions encounter challenges exchanging WSIs due to proprietary formats or inconsistent metadata — with some industry assessments estimating this affects up to 70% of organizations.[7][8] This not only slows clinical workflows but also limits multi-site research, vendor flexibility, and AI development.

"DICOM-WSI isn't about compliance — it's about breaking out of vendor lock-in and unlocking your data."

RUPESH MISHRA
HEALTH AI/ML SPECIALIST SOLUTIONS ARCHITECT, AMAZON WEB SERVICES (AWS)

Metadata Standards: The Hidden Backbone of Efficiency

High-quality metadata — the information about a slide — is often as important as the image itself. Fields like stain, magnification, specimen type, and encounter date determine whether a slide can be reliably retrieved, compared, or used in downstream workflows. When these fields are inconsistent or missing, teams spend time reconciling labels, restoring context, or manually repairing datasets.

Standardized metadata eliminates this friction. It enables fast case retrieval, consistent QA review, clean research handoffs, and reproducible AI development. In digital pathology and machine learning studies, inconsistent or incomplete metadata has been shown to increase data preparation time by 40–80%,[9][10] underscoring how much of the burden stems not from modeling but from fixing upstream variability.

FHIR and HL7: Connecting Pathology to the Rest of the Health System

FHIR (Fast Healthcare Interoperability Resources) and HL7 are standards for exchanging clinical data across systems. In pathology, they ensure that:

- WSIs can be linked reliably to clinical context
- pathology reports align with Electronic Health Record (EHR) systems
- downstream analytics can associate outcomes with image features
- multi-modal datasets (imaging + molecular + clinical) can be assembled for research

These standards make pathology data usable not just within the pathology department but across the entire care continuum.

"AI in pathology will only be as interoperable as the standards underneath it. DICOM-WSI isn't optional if we want scalable, trustworthy AI."

KEVIN SCHAP
DIRECTOR, CLINICAL INFORMATICS INITIATIVES, COLLEGE OF AMERICAN PATHOLOGISTS (CAP)

7 Journal of Pathology Informatics. Journal of Pathology Informatics. 2022;13. https://www.jpathinformatics.org/

8 Pure Storage. Digital pathology data solution brief. https://www.purestorage.com/content/dam/pdf/en/solution-briefs/sb-digital-pathology-data-pure.pdf

9 US Department of Health and Human Services, Office for Human Research Protections. Coded private information or specimens use in research: guidance. https://www.hhs.gov/ohrp/regulations-and-policy/guidance/research-involving-coded-private-information/index.html

10 US Food and Drug Administration. Medical devices; quality system regulation amendments (final rule). Federal Register. Published February 2, 2024. https://www.federalregister.gov/documents/2024/02/02/2024-01709/medical-devices-quality-system-regulation-amendments

Governance becomes meaningful only when it moves from written policy into daily decision-making. In leading pathology organizations, governance is a concrete set of habits, expectations, and structures that shape how data is generated, managed, shared, and protected. And when it functions well, it becomes invisible — workflows unfold smoothly, data is handled consistently, and teams know exactly how to proceed.

This section shows how mature organizations structure and run governance day to day.

05.01 A New Kind of Data Leadership

Modern pathology organizations are being asked to do two things at once: protect the integrity of the clinical record and enable high-quality secondary use of pathology data (WSI, LIS data, EHR context, molecular results). The practical answer is not more policies, it's clear decision rights, lightweight controls, and repeatable workflows that make the right thing the easy thing.

This is also a crucial moment to integrate practicing pathologists directly into data programs. As AI and biomarker development accelerate, the limiting factor is rarely computation, it's whether datasets reflect the diagnostic reality of the right cohort logic, the right ground truth, and the right workflow constraints.

For example, PIROuette Dx's PIRO model (Pathologist-Integrated Research Organization) embeds pathologists into the development loop to ask the "why" behind each request: why this cohort, why these stains, why these clinical variables, and what failure modes will matter in real diagnostic workflows? The resulting multimodal dataset is then, and only then, fit for purpose for AI development, biomarker discovery, diagnostics, and therapeutic/prognostic research. This integration reduces rework, increases label reliability, and shortens iteration cycles, enabling teams to move faster with higher-confidence data and a clearer line of sight to clinically meaningful and positively impactful patient outcomes.

A "new kind of data leadership" is an operating model that connects clinical authority to data execution. In practice, high-performing programs define three complementary roles:

Data Owner (Clinical): the accountable medical leader who defines clinical meaning, approves secondary use boundaries, and owns clinical risk decisions.

Data Steward (Operations/Informatics): the operational owner who standardizes metadata, maintains templates/logs, and ensures daily process adherence and traceability.

Data Custodian (IT/Security): the system owner who enforces access controls, audit logging, encryption, retention, and secure transfer mechanisms.

This structure aligns well with Quality Management (QM) expectations in medical laboratories (document control, responsibility, competence, and compliance management) while remaining compatible with medical-device Quality Management System (QMS) expectations when pathology AI workflows cross into regulated product development contexts (ISO 15189, ISO 13485, and FDA QMSR for devices).[11] [12] [13] [14]

Minimum governance artifacts that keep governance real and audit-ready without slowing workflow include:

- One-page Governance Charter (scope, decision rights, meeting cadence, escalation path)
- Controlled Data Dictionary and Permitted Metadata Schema per project
- De-identification and Release Checklist (spot-checks, logs)
- Custodian-Held Linkage Registry policy (if longitudinal linkage or re-pull is in scope)
- Deviation/incident workflow with a CAPA template and effectiveness checks

Done well, this minimal set of documents becomes the backbone for speed and teams do not debate ad hoc every time; they execute the same playbook with documented guardrails.

> "The laboratory is responsible for at least 70% of the data from which decisions to treat are made. Pathologists understand the data. To ensure quality and patient safety, pathologists must be front and center whenever any lab data is included in any AI model or algorithm."
>
> **M. E. DE BACA, MD**
> VICE PRESIDENT FOR MEDICAL AFFAIRS, SYSMEX AMERICA, GOVERNOR, COLLEGE OF AMERICAN PATHOLOGISTS (CAP), CHAIR, CAP COUNCIL ON INFORMATICS AND PATHOLOGY INNOVATION[15]

11 International Organization for Standardization. ISO 13485:2016 medical devices—quality management systems—requirements for regulatory purposes. https://www.iso.org/obp/ui/en/#iso:std:iso:13485:ed-3:v1:en

12 International Organization for Standardization. ISO 15189:2022 medical laboratories—requirements for quality and competence. https://www.iso.org/obp/ui/en/#iso:std:76677:en

13 US Food and Drug Administration. Medical devices; quality system regulation amendments (final rule). Federal Register. Published February 2, 2024. https://www.federalregister.gov/documents/2024/02/02/2024-01709/medical-devices-quality-system-regulation-amendments

14 US Food and Drug Administration. Quality management system regulation (QMSR): frequently asked questions. https://www.fda.gov/medical-devices/quality-management-system-regulation-qmsr/quality-management-system-regulation-frequently-asked-questions

15 Lennerz JK, Schneider J, Lauterbach A. How health data integrity can earn trust and advance health. Issues in Science and Technology. https://issues.org/health-data-integrity-lennerz-schneider-lauterbach/

05.02 Cross-Functional Governance in Action

Governance fails when it is either purely clinical (and ignores operational realities) or purely operational (and ignores clinical meaning and risk). Leading organizations use a small, cross-functional governance structure that makes routine decisions swiftly and escalates only truly novel or higher-risk decisions.

A workable two-tier model is:

- **Pathology Data Governance Council (monthly):** clinical owners, quality/compliance, privacy, security/IT, pathology informatics, and research operations. This council owns policy posture, exceptions, and risk tier definitions
- **Operational Workgroup (weekly):** Research Use Only (RUO) operational lead, histology lead, informatics/data analyst, project lead, and QA/compliance delegate. This working group owns execution, throughput, issue resolution, and release readiness

The council's output should be visible and auditable with a decision log (what was decided, by whom, and why) in addition to template updates that translate decisions into operational controls. Below is a minimum RACI (Responsible/Accountable/Consulted/Informed) chart that can be adopted per site:

TABLE 5
GOVERNANCE AND CONTROL RACI

	CLINICAL DATA OWNER	RUO OPS LEAD	DATA STEWARD	QUALITY / COMPLIANCE	PRIVACY	IT / SECURITY
PROJECT INTAKE APPROVAL AND PERMITTED METADATA SCHEMA	A	R	R	C	C	I
DE-IDENTIFICATION POSTURE (SAFE HARBOR VS EXPERT DETERMINATION) AND RELEASE CHECKLIST	A	R	R	A/R	C	C
RELEASE AUTHORIZATION (QA SIGN-OFF, PHI CHECKS, PACKING LIST/MANIFEST)	A	R	R	A	C	C
RE-PULL AUTHORIZATION (WITHIN IRB SCOPE) AND TICKET CLOSURE	C	R	R	A	C	I
INCIDENT/DEVIATION RESPONSE AND CAPA APPROVAL/ CLOSURE	C	R	C	A	C	C
SCHEMA CHANGE AND TRACEABILITY UPDATES	A	C	R	C	C	C

"For us, governance wasn't bureaucracy — it was clarity. Once roles were defined, our research turnaround times dropped immediately."

GIOVANNI LUJAN, MD, FCAP
PROFESSOR AND VICE CHAIR, PATHOLOGY INFORMATICS AND COMPUTATIONAL PATHOLOGY, THE OHIO STATE UNIVERSITY

05.03 Governance in Action

Across institutions, governance models vary, but the high-performing ones share a consistent design that constitutes intake, risk-tiering, controlled execution, release gates, and reconstructable traceability.

1) **Intake and scope control.** Every secondary use project begins with a Project Intake Form that documents purpose (RUO vs. clinical), disease/assay scope, permitted metadata schema, intended deliverables (WSI, blocks/slides, de-identified metadata, molecular results), and the IRB posture (covered protocol or reliance).
2) **Risk-tiering to protect speed.** Use three operational tiers:
   - **Tier 1 (Fast Lane)**
     a) De-identified dataset, no longitudinal linkage, minimum necessary metadata, no PHI movement outside custodian environment
     b) Tier 1 offers minimal governance burden and rapid cycle time
   - **Tier 2 (Controlled Linkage)**
     a) De-identified deliverables, but linkage is needed inside the custodian lab for re-pulls, quality reconciliation, or limited longitudinal refresh
     b) Tier 2 requires a custodian-held linkage registry, ticketing, and tighter audit logging
   - **Tier 3 (Highest Scrutiny)**
     a) Identifiable or limited dataset access within a controlled environment or workflows that may be used to support regulated clinical product development
     b) Tier 3 requires privacy/security review, stronger documentation, and potentially QMS-aligned evidence packages (ISO, QMSR)[16 17 18]
3) **Controlled execution.** Once tiered, execution follows a defined workflow: candidate identification—> pathologist screening/accept-reject—>redundancy/record-integrity check—> RUO block/slide creation with non-PHI identifiers—> QA recuts—> de-identified metadata export—> WSI scanning and spot-checks—> release authorization[19 20 21 22]
4) **Release gates (minimum necessary).** Release gates are not bureaucracy, they are the fastest way to ensure repeatability. A lean release gate set includes a QA sign-off for tissue adequacy, PHI check for files and embedded metadata, packing list reconciliation, and documented authorization by quality/compliance[23 24]
5) **Reconstructability.** At any point later, you should be able to answer: "What exactly was released, when, under what approvals, and from which source materials?" This is not only good governance, it is a practical expectation across quality systems and emerging AI regulatory frameworks emphasizing documentation, logging, and traceability.[25 26 27]

CHECKLIST
IS YOUR PATHOLOGY GOVERNANCE AI-READY?

	Do you have formalized data ownership and stewardship roles?
	Are metadata and WSI formats standardized (e.g., DICOM)?
	Are audit logs captured, reviewed, and operationalized?
	Is there a cross-functional governance committee with decision-making authority?
	Are research and clinical data requests routed through consistent, well-defined workflows?
	Are governance policies mapped to AI-readiness benchmarks (e.g., bias monitoring, traceability)?

16 Lennerz JK, Schneider J, Lauterbach A. How health data integrity can earn trust and advance health. Issues in Science and Technology. https://issues.org/health-data-integrity-lennerz-schneider-lauterbach/

17 US Food and Drug Administration. Medical devices; quality system regulation amendments (final rule). Federal Register. Published February 2, 2024. https://www.federalregister.gov/documents/2024/02/02/2024-01709/medical-devices-quality-system-regulation-amendments

18 US Food and Drug Administration. Quality management system regulation (QMSR): frequently asked questions. https://www.fda.gov/medical-devices/quality-management-system-regulation-qmsr/quality-management-system-regulation-frequently-asked-questions

19 College of American Pathologists. Retention of laboratory records and materials. Accessed March 6, 2026. https://www.cap.org/member-resources/practice-management/strategy-and-planning

20 Electronic Code of Federal Regulations. 42 CFR 493.1105—Standard: retention requirements. Accessed March 6, 2026. https://www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493/subpart-J/section-493.1105

21 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

22 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

23 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

24 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

25 International Medical Device Regulators Forum. Good machine learning practice for medical device development: guiding principles. Published 2025. https://www.imdrf.org/sites/default/files/2025-02/IMDRF_AIML%20WG_GMLP_N88%20Final.pdf

26 European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex. Published 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

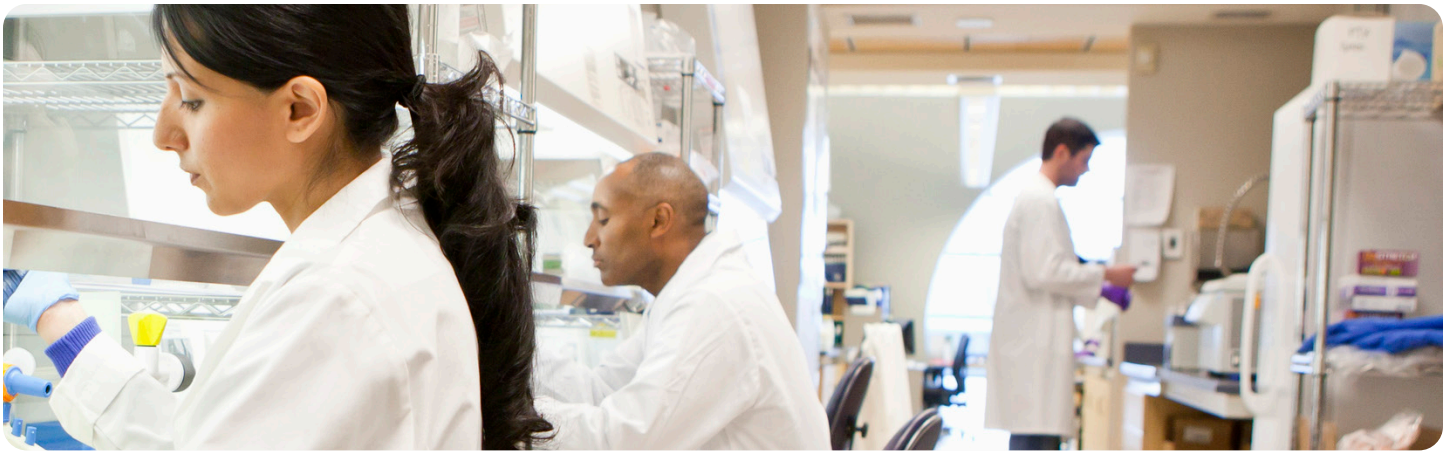
27 World Health Organization. Regulatory considerations on artificial intelligence for health. Published 2023. https://www.who.int/publications/i/item/9789240078871

05.04 Principles That Distinguish Leading Organizations

Organizations that scale pathology data ecosystems without compromising trust operationalize a small set of non-negotiable principles:

- **Clinical continuity first.** RUO workflows must never degrade clinical records. Redundancy, record integrity checks, and retention practices should align with CLIA and accreditation expectations. Many labs retain paraffin blocks far beyond federal minimum requirements to support patient care, re-review, and quality assurance.[28] [29] [30] Digital pathology complements this by creating durable, searchable digital archives that streamline access and reduce the operational burden of physical retrieval. As these archives mature, they enable AI-supported stratification workflows where H&E-based models can help prioritize cases for confirmatory testing, so scarce tissue is reserved for the assays influencing clinical or research decisions.
- **Minimum necessary by design.** Define a permitted metadata schema up front and enforce it technically and procedurally. When amendments are necessary, execute them through controlled change requests, not by creating exceptions. This standard mirrors the HIPAA de-identification posture and privacy by design expectations.[31] [32]
- **De-identification is a workflow, not an occasional action.** Safe Harbor vs. Expert Determination is a governance decision, but the daily protection happens through controls such as file naming, embedded metadata checks, label standards, secure transfer, and audit logs.[33] [34]
- **Custodian-held linkage is the default for longitudinal work.** If re-pulls or patient-level refresh is needed, keep the mapping in the custodian environment with least privilege access and ticketing. Do not distribute linkage tables externally.
- **Auditability is a product feature.** Build logs and checklists into the workflow so audits are simple, not heroic.
- **Automation where it matters.** Automate what is repetitive and error prone (PHI scanning, naming validation, checksum generation, release packet assembly) to protect both speed and safety.[35] [36]

> "The rapid digitalization of health care is affording us an unprecedented opportunity to solve complex medical problems and improve lives of people on a global scale."
>
> **JOHN HALAMKA MD, MS**
> PRESIDENT, MAYO CLINIC PLATFORM [37]



28 College of American Pathologists. Retention of laboratory records and materials. Accessed March 6, 2026. https://www.cap.org/member-resources/practice-management/strategy-and-planning

29 Electronic Code of Federal Regulations. 42 CFR 493.1105—Standard: retention requirements. Accessed March 6, 2026. https://www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493/subpart-J/section-493.1105

30 Electronic Code of Federal Regulations. 42 CFR Part 493—Laboratory requirements (Clinical Laboratory Improvement Amendments). Accessed March 6, 2026. https://www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493

31 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

32 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

33 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

34 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

35 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

36 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

37 Mayo Clinic News Network. Mayo Clinic launches its first platform initiative. https://newsnetwork.mayoclinic.org/discussion/mayo-clinic-launches-its-first-platform-initiative/

05.05 The Convergence of Clinical and Research Ecosystems

Pathology is uniquely positioned at the intersection of clinical diagnosis and translational research. The same physical specimen and interpretive context can support patient care, quality assurance, and discovery, but only if the operating model preserves the boundary conditions. There are three convergence pressures driving governance complexity:

1) **Digital derivatives behave like clinical records.** WSIs and their associated metadata can include patient identifiers if not intentionally controlled. A high functioning model treats WSI governance as part of the laboratory quality system (labeling, retention logic, access control, and documentation).[38] [39]
2) **Research workflows increasingly require clinical context.** RUO studies often require curated data with treatment context and molecular results. Without a "minimum necessary" schema and strong de-identification workflow, teams drift into uncontrolled data collection.
3) **AI governance expectations are tightening.** Even when current work is RUO, downstream use cases may include regulated AI. Emerging frameworks (EUAI, WHO regulatory considerations, IMDRF, and FDA expectations for AI/ML SaMD (Software as a Medical Device)) emphasize documentation, risk management, logging, and human oversight. These features can easily be employed proactively rather than retroactive damage control.[40] [41] [42]

The practical way forward is to design a single operating model, with strong traceability and controlled change, that can support RUO speed today while remaining upgradeable for regulated contexts tomorrow.

05.06 Best Practices for Modern Clinical Data Governance

This section translates governance theory into operational moves. The goal is to be simultaneously compliant and audit-ready, scalable across sites and teams, and fast enough to work in real pathology operations.

Each practice below is written as an operational pattern: what to implement, what evidence to capture, and how to keep it lightweight.

Four operational patterns distinguish leading programs: Centralized Authority, Risk-Based Automation, Operational AI Governance, Treating Data as a Strategic Asset.

Centralized, Cross-Functional Governance Authority

Implement a small governance council with clear decision rights and an explicit default path for routine requests. The council should not review every case, it should define the rules that make routine cases routable.

Minimum required outputs:

- Governance charter, membership roster, meeting cadence
- Decision log (including rationales for exceptions)
- Approved metadata schema per project (minimum necessary)
- Standard templates: intake form, screening sheet, packing list, PHI check record, incident report/CAPA[43] [44]

Evidence to retain:

- Decision log entries tied to project intake packets and release packets
- Training/competency records for staff executing RUO workflows (aligns with lab QMS expectations)

Risk-Based Automation

Automate what creates preventable risk and delays such as identifier checks, naming validation, embedded metadata screening, and release packet assembly.

Practical examples in digital pathology operations:

- Enforce file naming rules that prohibit PHI (no names, MRN, accession, full dates) and require Research ID and HashID
- Validate scanner configuration to suppress PHI in WSI headers/metadata and perform periodic spot checks on released WSI files[45] [46]
- Use automated PHI scanning for exported metadata files before release
- Generate checksums and a release manifest so downstream recipients can verify integrity

These are small automations that reduce rework and support the auditable controls expected across privacy and quality systems.

38 College of American Pathologists. Retention of laboratory records and materials. Accessed March 6, 2026. https://www.cap.org/member-resources/practice-management/strategy-and-planning

39 Electronic Code of Federal Regulations. 42 CFR 493.1105—Standard: retention requirements. Accessed March 6, 2026. https://www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493/subpart-J/section-493.1105

40 International Medical Device Regulators Forum. Good machine learning practice for medical device development: guiding principles. Published 2025. https://www.imdrf.org/sites/default/files/2025-02/IMDRF_AIML%20WG_GMLP_N88%20Final.pdf

41 European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex. Published 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

42 World Health Organization. Regulatory considerations on artificial intelligence for health. Published 2023. https://www.who.int/publications/i/item/9789240078871

43 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

44 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

45 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

46 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

Operational AI Governance

Even when the immediate project is RUO, governance should assume that high quality models may later be evaluated for clinical use. Build an AI-ready evidence trail without forcing a medical device program prematurely.

Minimum AI governance touchpoints to embed now:

- Intended use statement for each dataset/model (RUO vs. potential clinical use)
- Track data provenance, dataset versioning, and documented inclusion/exclusion criteria (traceability)
- Human oversight points (pathologist QA sign-off for tissue adequacy)
- Logging and documentation practices that can be expanded into technical documentation if needed (EU AI Act)[47]
- Alignment with widely recognized good practices (WHO regulatory considerations, IMDRF GMLP)[48 49]

Treating Data as a Strategic Asset — Not Just a Risk

Treating data as an asset does not mean maximal extraction, it means curating high quality, well documented datasets that can be reused safely.

Operationally, this looks like:

- A standardized data model that utilizes case-level Research ID, specimen/block/slide-level HashIDs, WSI identifiers, and metadata schema that supports multimodal linkage (WSI, LIS, EMR, and molecular results) without exposing PHI
- Reusable templates and controlled vocabularies for diagnoses, specimen types, assays, and outcomes
- A disciplined change-control process

From Governance in Theory to Governance in Practice

The fastest programs make governance executable by converting it into controls, workflow, and templates. In other words: "what must happen" (control), "who does it and when" (workflow), and "how we prove it happened" (template/evidence).

A minimum control set suitable for pathology RUO programs includes:

- Intake and IRB scope confirmation
- De-identification posture (Safe Harbor or Expert Determination) and enforcement[50 51]
- Custodian-held linkage registry controls for any longitudinal work
- Chain-of-custody and record integrity safeguards
- Traceability matrix and release gates with documented approvals
- Incident response, deviation handling, and CAPA with effectiveness checks

47 European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex. Published 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

48 International Medical Device Regulators Forum. Good machine learning practice for medical device development: guiding principles. Published 2025. https://www.imdrf.org/sites/default/files/2025-02/IMDRF_AIML%20WG_GMLP_N88%20Final.pdf

49 World Health Organization. Regulatory considerations on artificial intelligence for health. Published 2023. https://www.who.int/publications/i/item/9789240078871

50 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

51 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

Compliance is the foundation that shapes the ethical and legal boundaries within which pathology data must be created, stored, accessed, and shared. Governance builds upon this foundation, translating compliance obligations into daily operational behavior.

Understanding the compliance environment is essential for laboratory leaders, especially as data volumes grow, secondary uses expand, and AI applications emerge.

06.01 Federal Requirements: Protecting Privacy and Diagnostic Integrity

Three federal frameworks shape the compliance landscape for digital pathology. Building on the regulatory overview in Section 3, this section examines how these frameworks apply in practice.

HIPAA and HITECH govern how protected health information (PHI) is secured, transmitted, and audited. In a digital pathology context, PHI encompasses not only reports but also WSIs, metadata, annotations, case histories, and any additional artifacts linked to an identifiable patient. These standards apply equally across on-premises, hybrid, cloud, and distributed environments.

CLIA and CAP regulate diagnostic quality, laboratory operations, and record retention. As more laboratories transition to digital primary diagnosis, these expectations extend into digital imaging validation, scanner performance, and the completeness of digital archives.

The FDA plays an increasingly central role. Digital pathology viewers, WSI scanners, and computational tools that influence diagnosis or clinical decision-making may fall under FDA oversight. The FDA's principles of Good Machine Learning Practice (GMLP) emphasize transparency, reproducibility, data quality, and lifecycle monitoring.

Most pathology leaders are already familiar with these frameworks. What has changed is their scope of application in a digital environment. Three implications deserve particular attention:

First, PHI now lives in places it never did before. WSI file headers, embedded metadata, scanner-generated label images, and exported annotation files can all carry patient identifiers — often unintentionally. De-identification is a continuous workflow that must account for every digital artifact a case produces.

Second, record retention and integrity are more complex at digital scale. CLIA and CAP expectations for specimen retention, diagnostic completeness, and audit trails were designed for physical workflows. In digital environments, organizations must decide how long WSIs are retained, how archived images remain readable as formats evolve, and how to ensure that digital records are as durable and trustworthy as the glass slides and paraffin blocks they supplement.

Third, FDA oversight is expanding into software and algorithms. As AI-driven tools move from research into clinical decision support, organizations need governance structures that can demonstrate how data was curated, how models were validated, and how ongoing performance is monitored. The FDA's evolving guidance signals that traceability and documentation will be expected across the entire model lifecycle — not just at the point of regulatory submission.

OPERATIONAL INSERT (PIROUETTE DX)
DE-IDENTIFICATION, LINKAGE GOVERNANCE (HASHID), AND PHI-PREVENTION CONTROLS

This insert describes the minimum controls needed to produce de-identified pathology deliverables (RUO blocks/slides, WSI, and de-identified metadata) while preventing PHI leakage in file names and embedded metadata. It is not intended to be legal advice. Implement with local privacy/IRB policies and counsel:

1) Select and document a de-identification posture for each project

- Safe Harbor: remove the enumerated identifiers and ensure the remaining information cannot identify an individual[52] [53]
- Expert Determination: a qualified expert documents that re-identification risk is minimal given the context and controls[54] [55]

2) Use non-PHI identifiers by design

- Research ID: project case identifier used across all RUO deliverables (no PHI)
- HashID: a non-sequential, non-meaningful unique identifier generated from a cryptographically secure random generator (not derived from MRN, accession, or dates). HashIDs support cassette/slide/WSI naming while avoiding PHI
- Leading digital pathology platforms enforce de-identification by immediately assigning HashIDs at the point of image ingestion

3) Enforce PHI-prevention controls for WSI and exports

- File naming rules: prohibit patient name, MRN, accession, full DOB, full collection dates. Instead, require ResearchID and HashID for slide type/level/stain
- Embedded metadata checks: validate scanner/export settings, spot check WSI headers and associated metadata for identifiers before release (DICOM tags, proprietary header fields, and embedded slide label images can all be vectors)[56]
- Uniform labeling/printing controls: adopt standardized label content rules to reduce accidental inclusion of identifiers on RUO labels[57]

4) Evidence capture (minimum)

- PHI check record (who checked, what was checked, sample size, findings)[58] [59]
- Release manifest (what was released, versions, checksums if used)
- Release authorization sign-off (quality/compliance)

5) If longitudinal linkage or re-pull is required, use a custodian-held linkage registry model.

- Keep the mapping (ResearchID/HashID to clinical identifiers) inside the custodian environment, under Role-Based Access Control/Multifactor Authentication (RBAC/MFA) and audit logging. Linkage maps must not be distributed externally

52 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

53 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

54 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

55 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

56 DICOM Standards Committee. DICOM whole slide imaging (WSI). Accessed March 6, 2026. https://dicom.nema.org/dicom/dicomwsi/

57 College of American Pathologists. Uniform labeling of blocks and slides in surgical pathology. Accessed March 6, 2026. https://www.cap.org/protocols-and-guidelines/cap-guidelines/current-cap-guidelines/uniform-labeling-of-blocks-and-slides-in-surgical-pathology

58 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

59 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

OPERATIONAL INSERT (PIROUETTE DX)
INCIDENT RESPONSE + CAPA FOR DATA/SPECIMEN EVENTS

Pathology data operations necessitate incident response paths that are fast, documented, and corrective. If suspected PHI exposure, mislabeling, custody break, or improper export occurs.

- Contain: quarantine affected files/materials, suspend transfers, and preserve logs
- Assess and notify: activate privacy/compliance and determine whether breach criteria apply under applicable law/policy
- Correct: fix the immediate issue (reexport after metadata scrub, relabel, rescan)
- Prevent recurrence: root-cause analysis, CAPA plan, staff retraining, and effectiveness check. CAPA concepts are core to device/lab quality systems (FDA QMSR, ISO 13485, ISO 15189)[60] [61] [62] [63] [64]

06.02 Institutional Oversight: Ethics, Transparency, & Secondary Use

Within each organization, research and secondary use of pathology data are subject to additional layers of oversight. Institutional Review Boards (IRBs) determine when research use is permissible and whether informed consent or waivers are required.

Data Use Agreements (DUAs) establish the terms under which external parties may receive de-identified or limited datasets.

Regulatory definitions of de-identification — such as those outlined in 45 CFR §164.514 — guide how data must be stripped of identifiers before reuse.

These institutional controls are not simply legal mechanisms; they reinforce trust. When patients know their data is used transparently and ethically, research participation expands, and institutional credibility grows.

OPERATIONAL INSERT (PIROUETTE DX)
IRB-READINESS FRAMING (PRACTICAL, NOT LEGAL ADVICE)

For most pathology secondary use work, delays come from unclear scope and documentation, not from science. A practical IRB-readiness sequence is:

1. Confirm whether the work is human subjects research and what oversight applies (local IRB guidance and OHRP resources are a useful starting point).[65] In practice, consent requirements vary by use case. De-identified retrospective research that does not impact care is often exempt. Aggregate, population-level analyses using de-identified data may similarly not require individual consent, depending on institutional policy. Prospective engagement or re-contact for additional testing requires explicit, IRB-approved patient consent.
2. Lock the minimum necessary data elements (permitted metadata schema) and de-identification method.
3. Define linkage expectations explicitly (none vs. custodian-held linkage vs. limited dataset).
4. Operationalize consent/waiver language into workflows: intake —> execution —> release gates.
5. Maintain an IRB evidence packet that includes protocol/reliance confirmation, data dictionary, de-identification workflow, and release checklist aligned with documentation expectations in GCP contexts.[66]

60 Electronic Code of Federal Regulations. 21 CFR Part 820—Quality System Regulation. Accessed March 6, 2026. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-H/part-820

61 International Organization for Standardization. ISO 13485:2016 medical devices—quality management systems—requirements for regulatory purposes. https://www.iso.org/obp/ui/en/#iso:std:iso:13485:ed-3:v1:en

62 International Organization for Standardization. ISO 15189:2022 medical laboratories—requirements for quality and competence. https://www.iso.org/obp/ui/en/#iso:std:76677:en

63 US Food and Drug Administration. Medical devices; quality system regulation amendments (final rule). Federal Register. Published February 2, 2024. https://www.federalregister.gov/documents/2024/02/02/2024-01709/medical-devices-quality-system-regulation-amendments

64 US Food and Drug Administration. Quality management system regulation (QMSR): frequently asked questions. https://www.fda.gov/medical-devices/quality-management-system-regulation-qmsr/quality-management-system-regulation-frequently-asked-questions

65 US Department of Health and Human Services, Office for Human Research Protections. Coded private information or specimens use in research: guidance. https://www.hhs.gov/ohrp/regulations-and-policy/guidance/research-involving-coded-private-information/index.html

66 International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH E6(R3) guideline for good clinical practice. Published 2025. https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf

06.03 Global Standards: Cross-Border Expectations & AI Regulation

International regulations increasingly influence data governance, **even for U.S.-based laboratories**. The EU's GDPR has become a foundational privacy law shaping how organizations approach cross-border data transfers, vendor relationships, and multi-site research collaborations — even outside the EU.

The **EU AI Act** introduces additional, legally binding requirements for healthcare AI systems classified as "high-risk," including obligations related to risk management, documentation, transparency, human oversight, bias mitigation, and post-market monitoring.

Alongside these regulations, a growing number of professional bodies and standards organizations are publishing guidance on trustworthy and responsible AI. While these frameworks provide valuable direction, there is not yet global regulatory harmonization or consensus on a single standard.

Although U.S. regulatory policy continues to evolve, many vendors and international partners align their internal governance programs with GDPR, the EU AI Act, and emerging AI standards. As a result, American pathology departments engaged in global research, clinical trials, or multinational partnerships may encounter expectations shaped by regulatory and standards frameworks developed abroad.

These regulatory regimes and emerging standards may not apply directly to every pathology group, but they increasingly function as reference points for responsible data governance and AI oversight worldwide.

06.04 Compliance as the Baseline — Not the Endpoint

Compliance defines what organizations must do to protect patient privacy, ensure diagnostic integrity, and enable ethical secondary use. Governance governs how they do it. Together, they create a foundation that supports secure operations and paves the way for advanced digital pathology capabilities and AI readiness.

In practice, the most successful pathology organizations do not stop at satisfying regulatory requirements. They exceed them — using compliance as a foundation on which to build clarity, consistency, interoperability, and trust.

Operations are where governance holds—or breaks. This section outlines the day-to-day practices that protect data quality: QA controls, documentation, traceability, and cross-team alignment.

07.01 Embedding Quality Into Every Stage of the Workflow

Pathology data quality is earned upstream at specimen handling, block selection, labeling, scanning configuration, and metadata curation. The costly failures (unusable tissue, mislabeled slides, PHI leakage, or irreproducible datasets) are preventable with a small number of consistent checkpoints.[67] [68]

A minimum quality approach for RUO pathology workflows includes:

- **Tissue adequacy checkpoint:** pathologist screening and explicit accept/reject criteria documented on a screening sheet
- **Record integrity checkpoint:** redundancy check to ensure required diagnostic material remains in the clinical archive (especially for recent cases)[69] [70]
- **Labeling checkpoint:** RUO labels only use ResearchID/HashID. Follow uniform labeling principles to avoid accidental identifiers[71]
- **QA recut checkpoint:** every RUO block has an H&E QA slide reviewed and signed off before release.
- **Digital capture checkpoint:** scanner/export settings are validated to prevent PHI in filenames (embedded metadata) and periodic spot-checks are documented[72] [73]

These checkpoints are compatible with laboratory quality expectations and align with digital pathology best practices for validating WSI systems and workflows.

07.02 Operationalizing Governance Across Teams

Governance becomes a reality when it is built into how people work, not when it lives in a separate binder. The operational goal is to make compliant execution the default path across histology, informatics, quality/compliance, and clinical leadership.

A lean operational governance pattern looks like this:

- **Daily execution**
  - RUO ops lead and histology lead run the workflow
  - Informatics prepares candidate lists and de-identified exports
  - Pathologist performs screening and QA sign-off
  - Quality/compliance performs release authorization
  - **Weekly operations huddle**
    - Review throughput, exceptions, open tickets (including re-pull requests) and any deviations
- **Monthly governance council**
  - Review metrics (turnaround time, reject rates, PHI check findings), approve schema changes, and review any CAPA actions[74] [75]

Minimum required controls-as-workflow artifacts:

- Project Intake Form (scope, permitted metadata schema, deliverables)
- Candidate Screening Sheet (accept/reject and tissue adequacy)
- Chain-of-Custody Log (materials out/in and handlers)
- Redundancy Checklist (record integrity safeguarded)
- QA Sign-off Form (tissue confirmed)[76] [77]
- PHI Check Record and Release Authorization (de-identification verified)[78] [79]

Training and competency are vital to governance success. ISO 15189 expects defined responsibilities and competence for personnel performing tasks that affect quality. Mapping RUO roles to training checklists keeps execution consistent across sites.[80]

67 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

68 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

69 College of American Pathologists. Retention of laboratory records and materials. Accessed March 6, 2026. https://www.cap.org/member-resources/practice-management/strategy-and-planning

70 Electronic Code of Federal Regulations. 42 CFR 493.1105—Standard: retention requirements. Accessed March 6, 2026. https://www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493/subpart-J/section-493.1105

71 College of American Pathologists. Uniform labeling of blocks and slides in surgical pathology. Accessed March 6, 2026. https://www.cap.org/protocols-and-guidelines/cap-guidelines/current-cap-guidelines/uniform-labeling-of-blocks-and-slides-in-surgical-pathology

72 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

73 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

74 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

75 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

76 College of American Pathologists. Retention of laboratory records and materials. Accessed March 6, 2026. https://www.cap.org/member-resources/practice-management/strategy-and-planning

77 Electronic Code of Federal Regulations. 42 CFR 493.1105—Standard: retention requirements. Accessed March 6, 2026. https://www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493/subpart-J/section-493.1105

78 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

79 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

80 International Organization for Standardization. ISO 15189:2022 medical laboratories—requirements for quality and competence. https://www.iso.org/obp/ui/en/#iso:std:76677:en

07.03 Documenting Workflows and Maintaining Traceability

Traceability is not about creating paperwork, it's about being able to reconstruct what happened. In pathology data operations, that means linking: specimen —> block —> slide —> WSI —> exported metadata —> downstream dataset version.

A practical, minimum necessary traceability model uses:

- A stable Research ID at the case level (project and case)
- HashIDs at the artifact level (block/slide/WSI) to support naming and lineage without PHI
- A release manifest that records exactly what was released, when, and under what approvals

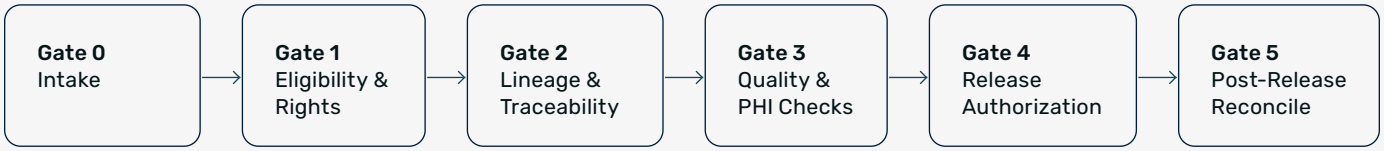
This level of traceability supports multiple overlapping expectations:

- Laboratory quality expectations for controlled records and handling (ISO 15189)[81]
- Clinical research expectations for documentation and data integrity (ICH GCP)[82]
- Emerging AI governance expectations emphasizing documentation, logging, and human oversight (WHO, IMDRF, EU AI Act)[83] [84]

PIROuette Dx recommends two "minimum viable" tools:

1) A Traceability Matrix that lists each artifact type, its identifier, required fields, and the evidence produced at each gate
2) Release Gates (Figure 2) that define the minimum checks required before a dataset or shipment is released (QA sign-off, PHI checks, packing list reconciliation, and authorization)[85] [86]

FIGURE 2
MINIMUM NECESSARY RELEASE GATES (RUO DATASETS AND ARTIFACTS)



"While data standards aren't recognized as sexy, neither is the stonework which makes up the foundations upon which our culture has been built. Only with strong data standards will we have data strong enough to use for AI. This includes but is not limited to lab data, digital data, and reporting standards."

M. E. DE BACA, MD
VICE PRESIDENT FOR MEDICAL AFFAIRS, SYSMEX AMERICA, GOVERNOR, COLLEGE OF AMERICAN PATHOLOGISTS (CAP), CHAIR, CAP COUNCIL ON INFORMATICS AND PATHOLOGY INNOVATION

81 International Organization for Standardization. ISO 15189:2022 medical laboratories—requirements for quality and competence. https://www.iso.org/obp/ui/en/#iso:std:76677:en

82 International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH E6(R3) guideline for good clinical practice. Published 2025. https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf

83 International Medical Device Regulators Forum. Good machine learning practice for medical device development: guiding principles. Published 2025. https://www.imdrf.org/sites/default/files/2025-02/IMDRF_AIML%20WG_GMLP_N88%20Final.pdf

84 European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex. Published 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

85 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

86 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

07.04 Supporting Scalable, Multi-Site Operations

Multi-site pathology operations introduce variation in accessioning, labeling, scanning, and metadata conventions. Scaling safely requires standardization at the interfaces (identifiers, metadata schema, and release evidence) while allowing local labs to keep their internal LIS processes.

Minimum design to support scaling:

- Standardize the de-identified identifier model (ResearchID/HashID) and apply it consistently across all RUO artifacts and file names.
- Use a common metadata schema (controlled vocabulary) and version it. Treat schema changes as change-controlled events.
- Prefer interoperability standards for WSI and metadata where feasible.[87] [88]
- Standardize release packet evidence so each site produces the same audit-ready deliverables (PHI check record, QA sign-off, packing list, manifest).[89] [90]

Consistency is the multiplier; it reduces downstream chaos and makes multi-site audits and IRB support dramatically easier.

07.05 Operational Readiness for AI: Building the Pipeline Before the Model

Operational readiness for AI starts before model training. It starts with dataset fitness from curated tissue, consistent labels, defined inclusion criteria, and documented provenance.

Minimum practices that enable AI work without over-engineering:

- Define intended use and analysis plan at intake (RUO discovery vs. eventual clinical evaluation)
- Maintain dataset versioning. "Dataset v1.0" should be reproducible from source artifacts and manifests
- Separate training, validation, and test cohorts, and document selection logic to reduce bias and PHI leakage
- Capture data shift risks early (stain variability, scanner variability, site effects) and document mitigations

These steps align with computational pathology best-practice themes emphasized by the Digital Pathology Association and broader AI governance guidance (EUAI, IMDRF).[91] [92]

07.06 What Operational Excellence Makes Possible

When governance and compliance are operationalized, pathology organizations can deliver high-trust multimodal datasets faster and with less friction through faster cohort builds (less rework, fewer downstream data corrections), cleaner data products (more usable WSIs, metadata, fewer artifacts), and lower privacy risk (fewer PHI leakage incidents, fewer downstream emergencies).

All this provides a foundation for responsible AI where datasets can support future validation and regulatory expectations. Operational excellence turns trust into throughput.

87 Digital Pathology Association; National Academy of Medicine. Procuring interoperability: achieving high-quality, connected, and person-centered care. Published 2019. https://nam.edu/wp-content/uploads/2019/08/Interop_508.pdf

88 DICOM Standards Committee. DICOM whole slide imaging (WSI). Accessed March 6, 2026. https://dicom.nema.org/dicom/dicomwsi/

89 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

90 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

91 International Medical Device Regulators Forum. Good machine learning practice for medical device development: guiding principles. Published 2025. https://www.imdrf.org/sites/default/files/2025-02/IMDRF_AIML%20WG_GMLP_N88%20Final.pdf

92 European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex. Published 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

AI is rapidly becoming part of the pathology landscape, offering the promise of improved diagnostic consistency, accelerated workflows, and deeper insight into complex disease patterns.

AI succeeds when the environment around it is designed for success. This section outlines how pathology organizations prepare for AI in practice: through data curation, governance maturity, ethical frameworks, and alignment with emerging regulatory expectations.

08.01 Data Quality as the Bedrock of AI Performance

Every AI model is only as good as the data that trains it. In pathology, this includes WSIs, metadata, annotations, case-level labels, and any derived measurements or QA attributes. Well-curated datasets — those with consistent metadata, high-quality images, clear provenance, and complete documentation — give models the best chance to perform reliably.

A strong data foundation is the single most important prerequisite for scaling AI initiatives, but Salesforce research found that 62% of IT leaders believe that their organization isn't yet equipped to fully leverage AI due to non-harmonized data systems.[93] Healthcare, however, may have a built-in advantage. For pathology — where HIPAA, CLIA, and CAP have long demanded rigorous data handling — this head start is real, but only if organizations convert compliance discipline into governance maturity.

Foundational practices such as scanner calibration, standardized metadata, rigorous quality assurance, and thorough documentation of dataset construction serve two parallel purposes: they support day-to-day diagnostic excellence, and they create the conditions necessary for training AI systems responsibly. If data quality is inconsistent, AI performance will be inconsistent. If data provenance is unclear, AI outputs cannot be trusted.

Industry reviews indicate that up to 80% of a data scientist's time is spent on data preparation[95] — underscoring that the real bottleneck in AI is not the algorithm but the dataset. Leading organizations recognize that building AI starts long before selecting a model. It begins with building a dataset users can stand behind.

> "Data governance should include deep, thoughtful consideration about the biases intrinsic in our society, as these permeate the data training sets if not consciously challenged."

M. E. DE BACA, MD
VICE PRESIDENT FOR MEDICAL AFFAIRS, SYSMEX AMERICA, GOVERNOR, COLLEGE OF AMERICAN PATHOLOGISTS (CAP), CHAIR, CAP COUNCIL ON INFORMATICS AND PATHOLOGY INNOVATION

> "Data governance is fundamentally the bedrock for ensuring patient safety. Healthcare organizations have already needed to clean and control their data. So, in a lot of ways, they're better positioned for AI than other industries."

THOMAS GODDEN,
AWS EXECUTIVE IN RESIDENCE,
FORMER CIO AT FOUNDATION MEDICINE[94]

93 Salesforce. 85% of IT leaders see AI boosting productivity, but data integration and overwhelmed teams hinder success. Connectivity Report announcement. 2024. https://www.salesforce.com/news/stories/connectivity-report-announcement-2024/

94 HealthTech Magazine. AI data governance in healthcare: what's new and what's changing? Published February 2025. https://healthtechmagazine.net/article/2025/02/ai-data-governance-in-healthcare-perfcon

95 Amazon Web Services. What is Data Preparation? Accessed March 6, 2026. https://aws.amazon.com/what-is/data-preparation/

08.02 Preparing Data Pipelines for Model Development

Once an organization has achieved consistency in data capture and governance, it can begin preparing the pipelines required for model training, validation, testing, and monitoring. These pipelines often include:

- structured processes for dataset assembly and de-identification,
- defined metadata taxonomies for labeling and stratifying cases,
- standardized annotation workflows,
- version-controlled storage environments for training and validation sets, and
- documentation that allows datasets to be reconstructed or audited long after the model is deployed.

These components form the "scaffolding" that makes AI development repeatable, auditable, and safe. Without such scaffolding, AI initiatives remain fragile prototypes rather than clinical-grade tools.

As AI becomes more deeply integrated into clinical and research workflows, traceability — knowing exactly which cases, annotations, scanners, and metadata contributed to a model — becomes essential. This is not just good practice; it is increasingly a regulatory expectation. According to a 2024 survey by Gartner, 63% of organizations either do not have — or are unsure if they have — the right data-management practices in place for AI.[96] This gap puts many AI initiatives at risk from the outset.

08.03 Governance Evolving into AI Governance

Traditional data governance provides clarity, stewardship, and quality control. AI governance builds on that foundation by introducing additional layers of oversight, including:

- monitoring for algorithmic drift,
- tracking model performance across patient subgroups,
- reviewing false positives and false negatives,
- updating models in controlled, documented cycles, and
- ensuring transparency in how model outputs are generated.

Organizations often expand their existing governance committees to include expertise from data science, biostatistics, clinical informatics, and quality management. This evolution reflects the reality that AI development is not separate from pathology operations — it is an extension of them.

A well-designed governance model makes AI safer by ensuring that decision-making authority, accountability, and documentation are already in place before models reach clinical deployment. In pathology, this oversight is not optional — any drift in model performance can directly impact diagnostic reliability.

"Anatomic Pathology is evolving from a predominantly morphologic discipline into an integrative, data-driven specialty, powered by advances in molecular diagnostics, digital pathology, and augmented intelligence. Strong data governance has become a strategic leadership priority, essential to ensuring that pathologists remain at the forefront of precision medicine."

MARILYN BUI, MD, PHD
SENIOR MEMBER & PROFESSOR OF PATHOLOGY, MOFFITT CANCER CENTER

96 Gartner. Lack of AI-ready data puts AI projects at risk. Press release. February 26, 2025. https://www.gartner.com/en/newsroom/press-releases/2025-02-26-lack-of-ai-ready-data-puts-ai-projects-at-risk

08.04 Ethical Frameworks and Risk Management

AI introduces ethical questions unlike those found in traditional diagnostic systems. Leaders must consider how models perform across diverse patient populations, how predictions are communicated to clinicians, and how accountability is maintained when both humans and algorithms contribute to a decision.

International organizations have begun to formalize best practices. In 2025, the International Medical Device Regulators Forum (IMDRF) built upon earlier FDA-, Health Canada–, and MHRA-led Good Machine Learning Practice (GMLP) principles, reinforcing expectations around data quality, reproducibility, lifecycle management, and ongoing performance monitoring.

While these frameworks differ in scope, they share a common message: AI requires continual vigilance. Pathology leaders play a central role in ensuring that models are deployed ethically, fairly, and responsibly.

08.05 Privacy-Preserving and Distributed Approaches

As interest in multi-site AI collaboration grows, so does the need for methods that protect patient privacy while enabling robust model development. Approaches such as federated learning, differential privacy, and secure enclaves allow institutions to contribute to AI models without sharing raw data.

Federated learning, for example, enables models to be trained across multiple institutions while patient data remains within each site's firewall. These approaches reduce risk, support compliance, and expand the diversity of data available for AI development — especially valuable in rare diseases or underrepresented populations. As pathology AI matures, these privacy-preserving technologies will likely become integral to collaborative research networks and commercial model development.

08.06 Emerging Standards and the Future Regulatory Landscape

Regulation for pathology AI is still evolving. Early FDA decisions for digital pathology software and computational tools provide valuable indicators, but the next generation of AI — including continuously learning models and multimodal systems — will demand new regulatory frameworks. The EU AI Act and similar emerging standards offer a preview: requirements for explainability, robustness, documentation, drift monitoring, and post-deployment reporting.

U.S. regulations may shift in similar ways, emphasizing lifecycle transparency and ongoing risk management. The institutions that prepare for these expectations now will be best positioned to adopt AI quickly and safely later.

OPERATIONALIZATION CALLOUT (PIROUETTE DX)
EVIDENCE PACK FOR AI-READY PATHOLOGY

Whether you operate under lab accreditation expectations (CAP/CLIA/ISO 15189) or anticipate regulated AI pathways (EU AI Act, MDR/IVDR, FDA QMSR, global good practices), the operational question remains the same: what evidence can you produce on demand? [97] [98] [99] [100] [101]

Minimum evidence artifacts to maintain continuously:

- Traceability matrix that links source materials—>WSI—>metadata—>dataset version—> release manifest
- Documented de-identification posture and PHI-prevention controls (including embedded metadata checks)[102] [103]
- Audit logs for linkage registry access and re-pull tickets (if longitudinal work is permitted)
- QA sign-off records and chain-of-custody logs
- Deviation/incident reports and CAPA records with effectiveness checks

These artifacts map naturally to quality management expectations (ISO 15189/ISO 13485/QMSR). They support documentation and logging expectations emphasized by emerging AI governance frameworks.[104] [105] [106] [107]

97 Electronic Code of Federal Regulations. 42 CFR Part 493—Laboratory requirements (Clinical Laboratory Improvement Amendments). Accessed March 6, 2026. https://www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493

98 International Organization for Standardization. ISO 15189:2022 medical laboratories—requirements for quality and competence. https://www.iso.org/obp/ui/en/#iso:std:76677:en

99 European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex. Published 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

100 US Food and Drug Administration. Medical devices; quality system regulation amendments (final rule). Federal Register. Published February 2, 2024. https://www.federalregister.gov/documents/2024/02/02/2024-01709/medical-devices-quality-system-regulation-amendments

101 US Food and Drug Administration. Quality management system regulation (QMSR): frequently asked questions. https://www.fda.gov/medical-devices/quality-management-system-regulation-qmsr/quality-management-system-regulation-frequently-asked-questions

102 Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514

103 US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

104 International Organization for Standardization. ISO 13485:2016 medical devices—quality management systems—requirements for regulatory purposes. https://www.iso.org/obp/ui/en/#iso:std:iso:13485:ed-3:v1:en

105 International Organization for Standardization. ISO 15189:2022 medical laboratories—requirements for quality and competence. https://www.iso.org/obp/ui/en/#iso:std:76677:en

106 US Food and Drug Administration. Medical devices; quality system regulation amendments (final rule). Federal Register. Published February 2, 2024. https://www.federalregister.gov/documents/2024/02/02/2024-01709/medical-devices-quality-system-regulation-amendments

107 US Food and Drug Administration. Quality management system regulation (QMSR): frequently asked questions. https://www.fda.gov/medical-devices/quality-management-system-regulation-qmsr/quality-management-system-regulation-frequently-asked-questions

08.07 What It Means to Be AI-Ready

AI-ready pathology departments share a common set of characteristics:

- They trust their data because it is governed, standardized, and high-quality.
- They maintain complete metadata, full provenance, and consistent QA practices.
- They support reproducible dataset creation and version-controlled pipelines.
- They have the governance structures necessary to oversee model development.
- They understand and prepare for emerging regulatory and ethical expectations.

In these organizations, AI readiness is not a future aspiration — it is a natural extension of their existing operating model.

> "AI-ready pathology is built on intentional governance, operational discipline, and cultural trust. Technology alone doesn't make you ready — foundations do."
>
> **MATTHEW HANNA, MD**
> VICE CHAIR OF PATHOLOGY INFORMATICS, ASSOCIATE PROFESSOR OF PATHOLOGY, UNIVERSITY OF PITTSBURGH MEDICAL CENTER

Building a Pathology Data Ecosystem You Can Trust

Pathology is entering a moment of transformation. Once anchored in physical media and local workflows, the discipline now operates within a dynamic data ecosystem that extends across scanners, systems, sites, clouds, collaborators, and increasingly, algorithms. This evolution presents extraordinary opportunities. It also demands a new level of rigor — not only in how images are captured or where data lives, but in how the entire lifecycle is governed, protected, and operationalized.

The throughline of this whitepaper is simple: trust is built long before AI enters the conversation. Trust is built when: metadata is consistent rather than improvised; governance is shared rather than siloed; compliance is understood not as a hurdle but as a foundation; operations reinforce quality at every step; and when organizations view their data not as exhaust from the diagnostic process, but as an institutional asset worthy of the same stewardship as any clinical instrument or medical device.

Leading pathology organizations are already demonstrating what this looks like. They establish governance models that connect clinical leadership, IT, compliance, operations, and research. They standardize formats, harmonize metadata, and embed quality into every workflow. They adopt infrastructure models aligned with long-term scalability. They implement access controls, audit trails, and documentation practices that strengthen both accountability and collaboration. Most importantly, they build cultures where data is handled with intention. With these foundations in place, AI becomes a safe extension of mature operations rather than a risky leap.

Organizations that invest in governance, compliance, quality operations, and ethical foresight position themselves to harness the full potential of digital tools and AI — not cautiously, not experimentally, but confidently and responsibly.

The opportunity ahead is significant. The organizations best prepared to seize it will be those that recognize what this paper has sought to make clear: data maturity is not ancillary to digital pathology or AI — it is the engine that makes progress possible. With the right foundation, pathology can realize a future where workflows are more consistent, insights more powerful, research more scalable, and patient care more precise.

That future is well within reach. The work begins now.

> "There are rare moments in history when technology, policy and urgency to change converge. I call that the 'perfect storm for innovation.'"
>
> **JOHN HALAMKA, MD, MS**
> PRESIDENT, MAYO CLINIC PLATFORM[108]

108 Mayo Clinic News Network. Platform revolution: curing more people, reaching more lives, anytime, anywhere. https://newsnetwork.mayoclinic.org/discussion/platform-revolution-curing-more-people-reaching-more-lives-anytime-anywhere/

1. College of American Pathologists. Retention of laboratory records and materials. Accessed March 6, 2026. https://www.cap.org/member-resources/practice-management/strategy-and-planning
2. DICOM Standards Committee. DICOM whole slide imaging (WSI). Accessed March 6, 2026. https://dicom.nema.org/dicom/dicomwsi/
3. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
4. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
5. US Department of Health and Human Services. Breach Notification Rule. https://www.hhs.gov/hipaa/for-professionals/breach-notification/index.html
6. Lee J, Takemaru L, Bappy DM, et al. Adaptive compression framework for giga-pixel whole slide images. Nat Commun. 2025. https://www.nature.com/articles/s41467-025-66889-0
7. Journal of Pathology Informatics. Journal of Pathology Informatics. 2022;13. https://www.jpathinformatics.org/
8. Pure Storage. Digital pathology data solution brief. https://www.purestorage.com/content/dam/pdf/en/solution-briefs/sb-digital-pathology-data-pure.pdf
9. US Department of Health and Human Services, Office for Human Research Protections. Coded private information or specimens use in research: guidance. https://www.hhs.gov/ohrp/regulations-and-policy/guidance/research-involving-coded-private-information/index.html
10. US Food and Drug Administration. Medical devices; quality system regulation amendments (final rule). Federal Register. Published February 2, 2024. https://www.federalregister.gov/documents/2024/02/02/2024-01709/medical-devices-quality-system-regulation-amendments
11. International Organization for Standardization. ISO 13485:2016 medical devices—quality management systems—requirements for regulatory purposes. https://www.iso.org/obp/ui/en/#iso:std:iso:13485:ed-3:v1:en
12. International Organization for Standardization. ISO 15189:2022 medical laboratories—requirements for quality and competence. https://www.iso.org/obp/ui/en/#iso:std:76677:en
13. US Food and Drug Administration. Medical devices; quality system regulation amendments (final rule). Federal Register. Published February 2, 2024. https://www.federalregister.gov/documents/2024/02/02/2024-01709/medical-devices-quality-system-regulation-amendments
14. US Food and Drug Administration. Quality management system regulation (QMSR): frequently asked questions. https://www.fda.gov/medical-devices/quality-management-system-regulation-qmsr/quality-management-system-regulation-frequently-asked-questions
15. Lennerz JK, Schneider J, Lauterbach A. How health data integrity can earn trust and advance health. Issues in Science and Technology. https://issues.org/health-data-integrity-lennerz-schneider-lauterbach/
16. Lennerz JK, Schneider J, Lauterbach A. How health data integrity can earn trust and advance health. Issues in Science and Technology. https://issues.org/health-data-integrity-lennerz-schneider-lauterbach/
17. US Food and Drug Administration. Medical devices; quality system regulation amendments (final rule). Federal Register. Published February 2, 2024. https://www.federalregister.gov/documents/2024/02/02/2024-01709/medical-devices-quality-system-regulation-amendments
18. US Food and Drug Administration. Quality management system regulation (QMSR): frequently asked questions. https://www.fda.gov/medical-devices/quality-management-system-regulation-qmsr/quality-management-system-regulation-frequently-asked-questions
19. College of American Pathologists. Retention of laboratory records and materials. Accessed March 6, 2026. https://www.cap.org/member-resources/practice-management/strategy-and-planning
20. Electronic Code of Federal Regulations. 42 CFR 493.1105—Standard: retention requirements. Accessed March 6, 2026. https://www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493/subpart-J/section-493.1105
21. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
22. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
23. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
24. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
25. International Medical Device Regulators Forum. Good machine learning practice for medical device development: guiding principles. Published 2025. https://www.imdrf.org/sites/default/files/2025-02/IMDRF_AIML%20WG_GMLP_N88%20Final.pdf
26. European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex. Published 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
27. World Health Organization. Regulatory considerations on artificial intelligence for health. Published 2023. https://www.who.int/publications/i/item/9789240078871
28. College of American Pathologists. Retention of laboratory records and materials. Accessed March 6, 2026. https://www.cap.org/member-resources/practice-management/strategy-and-planning
29. Electronic Code of Federal Regulations. 42 CFR 493.1105—Standard: retention requirements. Accessed March 6, 2026. https://www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493/subpart-J/section-493.1105

30. Electronic Code of Federal Regulations. 42 CFR Part 493—Laboratory requirements (Clinical Laboratory Improvement Amendments). Accessed March 6, 2026. https://www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493
31. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
32. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
33. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
34. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
35. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
36. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
37. Mayo Clinic News Network. Mayo Clinic launches its first platform initiative. https://newsnetwork.mayoclinic.org/discussion/mayo-clinic-launches-its-first-platform-initiative/
38. College of American Pathologists. Retention of laboratory records and materials. Accessed March 6, 2026. https://www.cap.org/member-resources/practice-management/strategy-and-planning
39. Electronic Code of Federal Regulations. 42 CFR 493.1105—Standard: retention requirements. Accessed March 6, 2026. https://www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493/subpart-J/section-493.1105
40. International Medical Device Regulators Forum. Good machine learning practice for medical device development: guiding principles. Published 2025. https://www.imdrf.org/sites/default/files/2025-02/IMDRF_AIML%20WG_GMLP_N88%20Final.pdf
41. European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex. Published 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
42. World Health Organization. Regulatory considerations on artificial intelligence for health. Published 2023. https://www.who.int/publications/i/item/9789240078871
43. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
44. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
45. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
46. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
47. European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex. Published 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
48. International Medical Device Regulators Forum. Good machine learning practice for medical device development: guiding principles. Published 2025. https://www.imdrf.org/sites/default/files/2025-02/IMDRF_AIML%20WG_GMLP_N88%20Final.pdf
49. World Health Organization. Regulatory considerations on artificial intelligence for health. Published 2023. https://www.who.int/publications/i/item/9789240078871
50. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
51. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
52. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
53. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
54. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
55. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
56. DICOM Standards Committee. DICOM whole slide imaging (WSI). Accessed March 6, 2026. https://dicom.nema.org/dicom/dicomwsi/

57. College of American Pathologists. Uniform labeling of blocks and slides in surgical pathology. Accessed March 6, 2026. https://www.cap.org/protocols-and-guidelines/cap-guidelines/current-cap-guidelines/uniform-labeling-of-blocks-and-slides-in-surgical-pathology
58. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
59. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
60. Electronic Code of Federal Regulations. 21 CFR Part 820—Quality System Regulation. Accessed March 6, 2026. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-H/part-820
61. International Organization for Standardization. ISO 13485:2016 medical devices—quality management systems—requirements for regulatory purposes. https://www.iso.org/obp/ui/en/#iso:std:iso:13485:ed-3:v1:en
62. International Organization for Standardization. ISO 15189:2022 medical laboratories—requirements for quality and competence. https://www.iso.org/obp/ui/en/#iso:std:76677:en
63. US Food and Drug Administration. Medical devices; quality system regulation amendments (final rule). Federal Register. Published February 2, 2024. https://www.federalregister.gov/documents/2024/02/02/2024-01709/medical-devices-quality-system-regulation-amendments
64. US Food and Drug Administration. Quality management system regulation (QMSR): frequently asked questions. https://www.fda.gov/medical-devices/quality-management-system-regulation-qmsr/quality-management-system-regulation-frequently-asked-questions
65. US Department of Health and Human Services, Office for Human Research Protections. Coded private information or specimens use in research: guidance. https://www.hhs.gov/ohrp/regulations-and-policy/guidance/research-involving-coded-private-information/index.html
66. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH E6(R3) guideline for good clinical practice. Published 2025. https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf
67. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
68. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
69. College of American Pathologists. Retention of laboratory records and materials. Accessed March 6, 2026. https://www.cap.org/member-resources/practice-management/strategy-and-planning
70. Electronic Code of Federal Regulations. 42 CFR 493.1105—Standard: retention requirements. Accessed March 6, 2026. https://www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493/subpart-J/section-493.1105
71. College of American Pathologists. Uniform labeling of blocks and slides in surgical pathology. Accessed March 6, 2026. https://www.cap.org/protocols-and-guidelines/cap-guidelines/current-cap-guidelines/uniform-labeling-of-blocks-and-slides-in-surgical-pathology
72. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
73. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
74. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
75. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
76. College of American Pathologists. Retention of laboratory records and materials. Accessed March 6, 2026. https://www.cap.org/member-resources/practice-management/strategy-and-planning
77. Electronic Code of Federal Regulations. 42 CFR 493.1105—Standard: retention requirements. Accessed March 6, 2026. https://www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493/subpart-J/section-493.1105
78. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
79. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
80. International Organization for Standardization. ISO 15189:2022 medical laboratories—requirements for quality and competence. https://www.iso.org/obp/ui/en/#iso:std:76677:en
81. International Organization for Standardization. ISO 15189:2022 medical laboratories—requirements for quality and competence. https://www.iso.org/obp/ui/en/#iso:std:76677:en
82. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH E6(R3) guideline for good clinical practice. Published 2025. https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf
83. International Medical Device Regulators Forum. Good machine learning practice for medical device development: guiding principles. Published 2025. https://www.imdrf.org/sites/default/files/2025-02/IMDRF_AIML%20WG_GMLP_N88%20Final.pdf
84. European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex. Published 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

85. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
86. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
87. Digital Pathology Association; National Academy of Medicine. Procuring interoperability: achieving high-quality, connected, and person-centered care. Published 2019. https://nam.edu/wp-content/uploads/2019/08/Interop_508.pdf
88. DICOM Standards Committee. DICOM whole slide imaging (WSI). Accessed March 6, 2026. https://dicom.nema.org/dicom/dicomwsi/
89. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
90. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
91. International Medical Device Regulators Forum. Good machine learning practice for medical device development: guiding principles. Published 2025. https://www.imdrf.org/sites/default/files/2025-02/IMDRF_AIML%20WG_GMLP_N88%20Final.pdf
92. European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex. Published 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
93. Salesforce. 85% of IT leaders see AI boosting productivity, but data integration and overwhelmed teams hinder success. Connectivity Report announcement. 2024. https://www.salesforce.com/news/stories/connectivity-report-announcement-2024/
94. HealthTech Magazine. AI data governance in healthcare: what's new and what's changing? Published February 2025. https://healthtechmagazine.net/article/2025/02/ai-data-governance-in-healthcare-perfcon
95. Amazon Web Services. What is Data Preparation? Accessed March 6, 2026. https://aws.amazon.com/what-is/data-preparation/
96. Gartner. Lack of AI-ready data puts AI projects at risk. Press release. February 26, 2025. https://www.gartner.com/en/newsroom/press-releases/2025-02-26-lack-of-ai-ready-data-puts-ai-projects-at-risk
97. Electronic Code of Federal Regulations. 42 CFR Part 493—Laboratory requirements (Clinical Laboratory Improvement Amendments). Accessed March 6, 2026. https://www.ecfr.gov/current/title-42/chapter-IV/subchapter-G/part-493
98. International Organization for Standardization. ISO 15189:2022 medical laboratories—requirements for quality and competence. https://www.iso.org/obp/ui/en/#iso:std:76677:en
99. European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex. Published 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
100. US Food and Drug Administration. Medical devices; quality system regulation amendments (final rule). Federal Register. Published February 2, 2024. https://www.federalregister.gov/documents/2024/02/02/2024-01709/medical-devices-quality-system-regulation-amendments
101. US Food and Drug Administration. Quality management system regulation (QMSR): frequently asked questions. https://www.fda.gov/medical-devices/quality-management-system-regulation-qmsr/quality-management-system-regulation-frequently-asked-questions
102. Electronic Code of Federal Regulations. 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Accessed March 6, 2026. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-164/subpart-E/section-164.514
103. US Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the HIPAA Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
104. International Organization for Standardization. ISO 13485:2016 medical devices—quality management systems—requirements for regulatory purposes. https://www.iso.org/obp/ui/en/#iso:std:iso:13485:ed-3:v1:en
105. International Organization for Standardization. ISO 15189:2022 medical laboratories—requirements for quality and competence. https://www.iso.org/obp/ui/en/#iso:std:76677:en
106. US Food and Drug Administration. Medical devices; quality system regulation amendments (final rule). Federal Register. Published February 2, 2024. https://www.federalregister.gov/documents/2024/02/02/2024-01709/medical-devices-quality-system-regulation-amendments
107. US Food and Drug Administration. Quality management system regulation (QMSR): frequently asked questions. https://www.fda.gov/medical-devices/quality-management-system-regulation-qmsr/quality-management-system-regulation-frequently-asked-questions
108. Mayo Clinic News Network. Platform revolution: curing more people, reaching more lives, anytime, anywhere. https://newsnetwork.mayoclinic.org/discussion/platform-revolution-curing-more-people-reaching-more-lives-anytime-anywhere/

DISCLAIMER

This whitepaper is provided for informational purposes only and does not constitute legal advice. The content reflects current technology capabilities, technical concepts, and industry practices as of the date of publication and may change without notice.

The information is provided as general guidance and should not be relied upon as a guarantee of specific outcomes or performance.

Readers are responsible for independently evaluating the information and determining its suitability for their particular use case.