LESSON 1.9: RECORD RETENTION AND AUDIT REPLAY

LESSON 1.9: SUPPLEMENTAL READING

Record retention requirements in regulated finance are often misunderstood as mandates to store vast amounts of information indefinitely. Firms approach retention as a technical problem:

- How much storage capacity is needed?
- How long must different record types be kept?
- What backup systems are required?

While these operational considerations matter, they miss the regulatory purpose of retention.

From a supervisory perspective, retention isn't about accumulating comprehensive archives. It's more about preserving the ability to reconstruct material decisions, demonstrate governance, and establish accountability when questions arise. Retention requirements exist to ensure that when regulators, auditors, or internal reviewers ask how a decision was made, the firm can provide contemporaneous evidence that answers that question credibly.

This distinction is critical. Firms may retain enormous volumes of data while still failing to preserve the records that matter most: the documentation of why data was selected, how it was interpreted, who approved its use, and what controls governed the process. Without these records, audit replay becomes impossible, and the firm's ability to demonstrate compliance is compromised regardless of how sound its practices may have been.

What Audit Replay Means in Practice

Audit replay refers to a firm's ability to reconstruct how a decision, analysis, or communication was produced at a specific point in time. It's the capacity to answer detailed, context-specific questions about past actions using contemporaneous records (documents, logs, approvals, and artifacts) created when the decisions were made, not reconstructed afterward.

When regulators conduct examinations or respond to complaints, they routinely ask firms to replay decisions:

What data was used? Not just what datasets exist now, but what specific data informed this decision at that time. If the data has since been updated, corrected, or deleted, can the firm reproduce the version that was actually used?

Why was it used? What business purpose, analytical need, or regulatory requirement justified using this data? What alternatives were considered? Why was this dataset deemed appropriate?

How was it interpreted or transformed? What cleaning, aggregation, derivation, or enrichment steps were applied? What assumptions governed these transformations? Were there judgment calls about handling missing values, outliers, or inconsistencies?

Who reviewed or approved its use? What oversight process was followed? Who had responsibility for validating that the data was fit for purpose? Is there a record of their review and approval?

What alternatives or limitations were considered? Were risks, biases, or sampling limitations identified? Were alternative approaches evaluated? How were trade-offs resolved?

Audit replay typically doesn't require perfect reconstruction of every interaction. It requires sufficient records to establish a coherent and credible narrative. Regulators understand that not every conversation, email, or minor decision can be preserved. But material decisions (i.e., those that shaped outcomes, influenced client treatment, or reflected supervisory judgment) must be traceable through contemporaneous documentation.

When firms can't replay decisions, regulators must assume that governance was informal, oversight was absent, or controls were inadequate. The inability to reconstruct is itself evidence of weak governance.

Retention Must Align with the Data Lifecycle

A common and critical weakness in record retention is retaining only final outputs while upstream data decisions and transformations are lost. Firms diligently preserve completed reports, client communications, regulatory filings, and model outputs.

But they fail to retain the records that explain how those outputs were produced upstream, such as the data selection decisions, transformation logic, interpretive assumptions, and supervisory approvals that preceded the final deliverable.

Regulators assess recordkeeping **across the full data lifecycle**, not just at endpoints. They want to understand not only what was communicated or decided, but how the firm arrived at that point. If the firm can produce a polished final report but can't explain where the underlying data came from, how it was validated, or who approved its use, the report itself becomes suspect.

Records that must be retained to support audit replay include:

Data selection and sourcing records: Documentation of why specific datasets were chosen, what alternatives were considered, and what criteria determined appropriateness. If third-party data was licensed, retention should include the licensing terms and any constraints on use.

Transformation and derivation logs: Records of what steps were applied to clean, aggregate, or derive new data elements. This includes the logic governing transformations, decisions about handling anomalies, and validations that outputs were correct.

Interpretive frameworks and assumptions: Documentation of how categories were defined, what thresholds were set, and what assumptions underlay scoring, ranking, or classification schemes. These frameworks often evolve over time, and firms must be able to show what framework was in effect at a given moment.

Approval and oversight records: Evidence that designated supervisors reviewed data use, validated appropriateness, and approved the application. This includes records of escalations, exceptions, or conditions placed on approvals.

Purpose definitions and changes: Documentation of the original purpose for which data was collected or used, and records of any subsequent changes to that purpose. If data were **repurposed**, there should be evidence that the reuse was reviewed and approved.

Monitoring and validation results: For data used in ongoing processes or automated systems, records of how controls performed over time. Were issues detected? How were they addressed? What trends or patterns emerged?

When retention policies are misaligned with how data actually flows through the organization, audit replay becomes difficult or impossible. For example, if formal retention policies apply only to published reports but not to the spreadsheets, scripts, or databases used to prepare them, critical evidence is lost.

Informal Systems Create Replay Gaps

Many data-related decisions occur in informal systems (think spreadsheets stored on personal drives, ad hoc SQL queries run in sandbox environments, email threads discussing analytical choices, collaboration tools where teams coordinate work). These systems are operationally efficient and support agile workflows, but they create significant replay gaps.

Informal systems often fall outside formal retention and archiving frameworks. They're subject to deletion when employees leave, expiration when storage quotas are exceeded, or loss when systems are decommissioned. Even when informal records persist, they

may be scattered across multiple locations, formats, and accounts, making them difficult to retrieve or piece together during an examination.

When regulators ask firms to replay a decision and the critical records exist only in fragmented or transient informal systems, the firm may be unable to produce a coherent narrative. Emails may be incomplete. Spreadsheets may have been overwritten. Scripts may have been deleted. Collaboration threads may have expired.

Regulators generally treat such gaps as **governance failures**, even when the underlying decisions were reasonable and well-intentioned.

The Interplay Between Retention and Replay

Retention and replay are distinct but interdependent concepts. **Retention** ensures that records are preserved for the required period. **Replay** ensures that those records can be assembled, interpreted, and presented in a way that reconstructs the decision context.

A firm may have strong retention practices (archiving records according to policy, preserving data for the required period, maintaining backup systems) but still fail at audit replay if:

- Records are preserved but not indexed or cataloged, making them difficult to locate when needed
- Records are retained in formats that become obsolete or inaccessible over time
- Records are fragmented across systems without clear linkages showing how they relate to each other
- Records exist but lack sufficient context to be interpretable—undocumented acronyms, missing explanations, references to decisions made elsewhere

Effective replay requires not just preservation, but **organization, accessibility, and context,** and this is why retention policies should be designed with replay in mind, not just compliance with minimum storage periods. The question is not only "How long must we keep this?" but also "Can we use this to reconstruct decisions if asked?"

Why This Matters Before Analytics or AI

Analytics platforms, machine learning models, and AI-driven systems accelerate decision-making and obscure intermediate steps. A model may generate thousands of predictions, classifications, or recommendations in seconds, with minimal human involvement at each step. The efficiency is valuable, but it creates significant replay challenges.

When regulators examine AI-assisted or automated workflows, they ask:

- What data was used to train or calibrate the system?

- How was that data selected, transformed, and validated?
- What assumptions were embedded in the model's design?
- Who approved the use of this data for this purpose?
- How has the model's performance been monitored over time?
- When anomalies or errors were detected, what actions were taken?

Firms that haven't designed retention and replay capabilities into their automated systems struggle to answer these questions. The system may function correctly, but the audit trail is incomplete or nonexistent. Records of training data may have been discarded. Transformation logic may exist only in undocumented code. Approval records may be scattered across email threads. Monitoring results may not have been preserved.

Regulators don't accept system complexity as an excuse for incomplete records. In fact, they often expect firms using advanced tools to demonstrate stronger retention and replay capabilities, not weaker ones. The opacity and speed of automation heighten the need for clear, retrievable evidence of governance and oversight.

Many firms assume replay will be possible later because systems "log activity." For example, LLM workflows often lose critical context unless replay is designed intentionally.

Common failure points include:

Ephemeral inputs: Prompts, context windows, or retrieved documents aren't retained, making it impossible to know what the model actually saw.
Model drift: The underlying model is updated or replaced, with no record of which version produced which outputs.
Undocumented orchestration: Decision logic exists across scripts, APIs, or workflow tools without a unified record of how components interacted.
Human judgment gaps: Reviews or approvals occur verbally or in chat, leaving no durable evidence of oversight.

When regulators ask how a particular output was produced, these gaps prevent firms from reconstructing the narrative of decision-making. The result isn't a technical failure, but a governance failure.

The solution

The best practice is to design retention and audit replay into workflows before automation scales to preserve accountability:

- Training data, validation data, and test data should be versioned and archived with metadata explaining their purpose and provenance

- Transformation scripts, feature engineering logic, and preprocessing steps should be documented and retained in version-controlled repositories
- Model approval records, validation reports, and monitoring results should be preserved in formal governance systems
- Incidents, errors, and corrective actions should be logged and retained with sufficient detail to support reconstruction

This upfront investment ensures that when questions arise (in examinations, audits, or incident investigations), the firm can demonstrate not just what the system did, but why it was designed that way, who approved it, and how it was supervised.

Conclusion

Record retention and audit replay aren't administrative requirements. They're the mechanisms that allow firms to demonstrate accountability, reconstruct decisions, and defend their practices under scrutiny. Retention ensures that evidence is preserved. Replay ensures that evidence can be assembled into a coherent narrative that answers regulators' questions.

Firms that retain only final outputs while losing upstream data decisions, transformations, and approvals compromise their ability to demonstrate compliance, regardless of how sound their practices may be. Firms that rely on informal systems for material decisions create replay gaps that regulators treat as governance failures.

Certain AI and analytics tools introduce additional replay requirements. Where models generate outputs based on prompts, retrieved context, or evolving configurations, firms may need to retain more than just results. Prompt logic, model versions, configuration settings, and access scopes can become part of the decision record, because they shape how outputs are produced and interpreted. Without these elements, firms may be unable to explain why a system behaved as it did at a specific point in time.

Before introducing analytics, automation, or AI, firms must design retention and replay capabilities into their data workflows. This foundation ensures that when systems scale, accountability remains traceable, decisions remain reconstructible, and governance remains demonstrable.