DATA ARCHITECTURE HEALTH ASSESSMENT

GUENIX

DATA ARCHITECTURE HEALTH ASSESSMENT

Before deploying AI or automation systems, institutions must assess their data architecture to identify fragmentation that will transform deployment into institutional liability. This checklist guides you through evaluating the three critical types of data fragmentation that break algorithmic models and cause automated systems to make high-velocity decisions based on incomplete or incoherent information.

Checklist

1. **Semantic Fragmentation Assessment**
   - ☐ **Map departmental definitions of core institutional terms**
     Identify how different departments define key terms like "beneficiary," "risk," "household," or "case." Semantic fragmentation occurs when the same label means different things across units, causing AI models to calculate correlations between fundamentally different concepts. This assessment reveals whether your data architecture supports coherent analysis or generates precise calculations based on linguistic misunderstandings.
   - ☐ **Document variations in data classification systems across units**
     Examine how regional offices, field teams, and headquarters classify the same types of information (eligibility criteria, risk categories, service levels). Inconsistent classification systems prevent algorithmic

models from learning meaningful patterns and force systems to treat conceptually identical cases as different categories, undermining predictive accuracy.

- ☐ **Identify conflicting business rules encoded in different systems**
  Review the logic embedded in separate departmental databases to find where the same process (eligibility determination, resource allocation) operates under different rules. When AI attempts to learn from this data, it will encode these conflicts as valid patterns, systematically replicating institutional incoherence at scale.
- ☐ **Audit data dictionaries and metadata standards**
  Determine whether standardized data dictionaries exist and, critically, whether they are actually used in practice. The gap between documented standards and operational reality reveals the degree to which semantic alignment is aspirational versus operational, directly impacting whether AI can function across departmental boundaries.

2. Temporal Fragmentation Assessment

- ☐ **Analyze the lag between events and data entry across processes**
  Map the time delay between when activities occur in the field and when they are logged in systems. Temporal fragmentation distorts causality—AI models analyzing this data will identify activity patterns based on administrative convenience rather than actual operational triggers, leading to systematically flawed predictions.
- ☐ **Identify timestamp reliability across critical datasets**
  Examine whether timestamps reflect when events happened or when staff had capacity to log them. In crisis response or field operations, this distinction determines whether your data reveals genuine patterns

or merely documents the rhythm of bureaucratic processing, rendering time-series analysis meaningless.

- ☐ **Document retroactive data entry practices**
  Assess how frequently records are entered days or weeks after events occur, and whether these delays are random or systematic (e.g., consistently delayed reporting from specific regions). Systematic delays create phantom correlations that AI will learn as causal relationships, baking temporal distortion into automated decision-making.
- ☐ **Evaluate event sequencing integrity in linked records**
  Test whether your systems preserve accurate chronological ordering when multiple events affect the same case or beneficiary. If system limitations or data migration have scrambled sequences, AI models cannot learn genuine cause-effect relationships, transforming predictive analytics into sophisticated guesswork.

3. **Structural Fragmentation Assessment**

- ☐ **Map the physical location of logically related data**
  Identify all systems storing information about the same entities (individuals, households, projects, partners) and documents which are connected versus isolated. Structural fragmentation occurs when records that should be analyzed together cannot be linked, forcing AI to make predictions based on thin slices of reality while missing critical context.
- ☐ **Audit unique identifier consistency across databases**
  Determine whether common identifiers (beneficiary IDs, case numbers, partner codes) exist across systems and whether they are used consistently. Without reliable linking mechanisms, even organizationally adjacent data becomes operationally unreachable to algorithmic

systems, preventing feature-rich analysis that humans would instinctively perform.

- ☐ **Assess data access barriers between departments**
  Examine the political, technical, and policy obstacles preventing data sharing across units. Structural fragmentation is often a political outcome reflecting organizational power dynamics—departments retain data to maintain autonomy. No technical solution resolves this until governance establishes trust frameworks for sharing without loss of control.
- ☐ **Identify shadow databases and unofficial record systems**
  Discover where staff maintain parallel spreadsheets, personal databases, or informal tracking systems because official systems are inadequate. These shadow systems contain institutional knowledge that AI cannot access, creating a systematic gap between the complete picture humans reference and the partial view algorithms see.

4. Data Governance Readiness

- ☐ **Document data ownership and stewardship responsibilities**
  Clarify which departments own which datasets and who bears responsibility for data quality, updates, and access decisions. Fragmentation persists when ownership is ambiguous, making it politically risky for any unit to champion integration without clear authority and protection.
- ☐ **Create data sharing agreements that protect departmental concerns**
  Develop governance frameworks that allow departments to share data without losing control, addressing legitimate concerns about nuance, confidentiality, and performance

accountability. The trust architecture enabling sharing must precede the technical infrastructure making it possible.

- ☐ **Establish a human override protocol**
  Define formally which categories of algorithmic decision require mandatory human review before execution. For each category, identify the role that holds override authority, the conditions under which an override is permissible, and the documentation required to protect the officer from adverse audit outcomes. An institution that deploys an automated decision system without this protocol has not kept a human in the loop. It has created the appearance of one.

- ☐ **Document the escalation pathway**
  Establish a formal escalation route for cases that fall outside the model's parameters or produce outputs that contradict field reality. Define who receives escalated cases, within what timeframe, and under what authority a decision must be made. Without a documented pathway, ambiguous cases accumulate in processing queues and the override mechanism exists technically while remaining politically disconnected.

5. Auditability Readiness

- ☐ **Version-control all governance artefacts**
  Confirm that data dictionaries, decision rights protocols, override logs, and metadata records are versioned, dated, and attributed to a named authority. Each artefact must be retrievable in the exact form it existed at the time of any given decision. Without version control, the institution cannot reconstruct the conditions under which a decision was made and accountability becomes unverifiable.

- ☐ **Establish and maintain an audit trail**
  Verify that the system produces a retrievable log of every decision it influences, including the data inputs used, the model version running at the time, and the output produced. This log must be accessible to internal audit functions and, where applicable, to external regulators, donors, or oversight bodies. The absence of a retrievable audit trail does not reduce institutional liability. It increases it.
- ☐ **Designate a named audit authority**
  Identify the role, whether a Data Governance Committee, a Chief Data Officer, or an equivalent function, responsible for conducting formal auditability reviews at defined intervals. Confirm that this role has the mandate, the access, and the resources to conduct reviews without requiring prior approval from the teams being reviewed.
- ☐ **Confirm explainability of decisions**
  Test whether the system can produce, for any output it generates, an explanation accessible to the individual affected. A system whose logic cannot be explained to the people it affects cannot be governed by the institution that deploys it.

6. Procurement Governance

- ☐ **Verify data provenance disclosure**
  Confirm that the vendor has documented and disclosed which datasets were used to train the model, the temporal range of those datasets, the definitions applied to core variables, and any known limitations or biases identified during model development. This documentation must be provided before contract signature, not after deployment.

- ☐ **Secure the right of inspection**
  Confirm that the contract grants the institution the right to audit the model's training data, architecture, and performance logs at any point during the contract period, without requiring prior notice or vendor cooperation beyond providing access. A vendor that refuses this clause cannot be held accountable.
- ☐ **Confirm explicit liability allocation**
  Verify that the contract specifies which party bears responsibility for decisions made on the basis of model outputs, particularly where the model's recommendation is followed without human override. Silence on this question concentrates liability on the institution.
- ☐ **Establish model update notification requirements**
  Confirm that the vendor is contractually required to notify the institution of any material change to the model before that change is deployed in a live environment, and that the institution retains the right to refuse or delay an update pending its own governance assessment.
- ☐ **Verify data sovereignty clauses**
  Confirm that the contract specifies where data is stored, under which legal jurisdiction, and under what conditions the vendor may use institutional data for purposes beyond the contracted service.