FEDERAL AI & DATA PRACTICE · RESEARCH REPORT

The Federal AI-Readiness Report

A five-dimension model for measuring whether a federal data estate can actually carry the AI workloads agencies are procuring — and a staged sequence for closing the gap.

FCI Advisory · McLean, Virginia

2026

Table of contents

Executive summary

Federal AI policy assumes agencies operate with AI-ready data. They do not. The procurement cycles in market right now are landing on data estates that have never been comprehensively audited for quality, classified for retrieval, or instrumented for lineage. The gap between policy assumption and operational reality is one of the largest hidden line items in federal AI program economics — and it is paid either upfront, before the pilot launches, or in remediation after the pilot fails to ship.

5	~62%	~4%	2→3
non-substitutable dimensions of AI-ready data — the composite is the **floor**, not the average	of agencies with AI in market sit below the "approaching" readiness threshold[4]	are fully operational AI-ready against retrieval workloads[4]	the fundable maturity move for most agencies this budget cycle

This report sets out the framework FCI Advisory uses to measure federal data readiness for AI and to sequence the work of closing the gap. Its argument is in five parts:

- **"AI-ready" has a specific operational meaning.** Federal data is ready only when five dimensions — completeness, accuracy, classification, lineage, and accessibility — hold *simultaneously* against the workloads the agency intends to run. The five do not average; the weakest dimension is the binding constraint.
- **The federal estate is not where the policy framework assumes it is.** The bulk of agencies with AI workloads in market or pilot sit in the lower-middle maturity bands — enough discipline to catalog their data, not enough to pass an AI retrieval workload.
- **The cost of the gap is real, and the only choice is timing.** Remediation done before deployment is bounded project work; remediation done after a stalled pilot is done with the lights on, with active users and accumulated audit obligations, at materially higher total cost.
- **Readiness is staged, not binary.** It moves through four stages — ad hoc, cataloged, governed, AI-ready. For most agencies, the realistic and fundable next move is Stage 2 to Stage 3 inside this budget cycle.
- **The sequence is the lever.** Inverting the standard order — readiness assessment and targeted remediation *before* the AI build, rather than after — lowers total program cost and raises the pilot success rate using the same components.

> *"Federal AI policy assumes the data is ready. The procurement cycles assume the data is ready. The model vendors assume the data is ready. The data is not ready. The cost of closing that gap is real, and the only choice is whether to pay it before the pilot or after."*

A note on figures. The directional patterns, distribution shares, and stage percentages in this report are drawn from FCI Advisory's engagement base across federal data-quality and AI-readiness programs, FY24-Q4 through FY26-Q1.[4] They are illustrative of the directional pattern, not point-precise survey extracts, and are presented to make the shape of the problem visible — not as audited statistics. Where a figure would imply a precision we cannot defend, we have framed it as directional or omitted it.

Why this report, and why now

Federal AI procurement has moved from pilot to program. Agencies are no longer asking whether to deploy AI against mission workflows; they are awarding multi-year contracts to do it. The policy scaffolding is in place — OMB Memorandum M-24-10 sets the governance and risk-management expectations, and the NIST AI Risk Management Framework supplies the functional vocabulary agencies are expected to operate against.[12] Both documents are well-constructed. Both make the same quiet assumption: that the data an agency intends to point its AI at is ready to be pointed at.

That assumption is the subject of this report. It is not an academic concern. The single most common reason a federal AI pilot fails to graduate to production is not the model, the cloud, or the policy — it is that the data underneath could not carry the workload, and nobody measured that before the build began. The cost of discovering this in production, rather than in planning, is the largest avoidable line item in federal AI program economics.

The window for getting ahead of it is the current budget cycle. The agencies that fund a readiness assessment and a scoped remediation now will deploy on data that can carry the workload. The agencies that defer will deploy on data that cannot, discover the gap when the pilot stalls, and remediate under worse conditions and worse economics. This report is written for the federal CIO, chief data officer, and AI program lead who would rather pay for the gap deliberately than discover it.

Who this report is for

- **Federal CIOs and CDOs** deciding what to fund in the AI program plan, and in what order.
- **AI program and product leads** scoping a pilot and deciding the go / no-go criteria for moving to production.
- **Acquisition and program-management staff** writing evaluation criteria for AI procurements, who need to score the data layer rather than only the model layer.

- **Records and governance officers** who will inherit the AI-generated outputs and the classification and lineage discipline that makes them manageable.

What the policy framework assumes — and leaves unsaid

M-24-10 and the NIST AI RMF both treat data quality, lineage, and representativeness as inputs to responsible AI use, not as deliverables an agency must first produce. The frameworks tell an agency that its data must be fit for purpose; they do not tell it how to measure whether it is, or what to do when it is not. That measurement and that remediation are exactly the gap this report addresses. The five-dimension model is, in effect, the operational instrument the policy frameworks presuppose but do not provide.

01 What "AI-ready data" actually means

The phrase "AI-ready data" gets used loosely across federal AI conversations. Stripped of the marketing layer, it has a specific operational meaning. Federal data is AI-ready when it satisfies five dimensions simultaneously, measured against the workloads the agency intends to run.

The five dimensions are not new. Federal records officers, database administrators, and data stewards have been measuring each of them for decades, with decades of definition behind them in records-management and database-administration practice. What is new is the requirement that all five hold *at the same time*, against AI retrieval workloads, with no opportunity to compensate for a weak dimension by being strong elsewhere.

The five dimensions

- **Completeness.** No missing fields or records. Retrieval cannot fill gaps the source does not contain; incomplete data produces incomplete results, which the model fills in with confident hallucination.
- **Accuracy.** Values reflect operational reality. Wrong inputs produce accurate-sounding wrong answers — confidently wrong outputs the agent has no way to flag.
- **Classification.** Labels and metadata enable targeted retrieval. Unclassified content cannot be queried; poorly classified data produces queries the agent cannot target.
- **Lineage.** Provenance is traceable. Without lineage, the agent's audit trail is unverifiable — a fatal problem in a federal environment where every action must be

defensible.

- **Accessibility.** The data is reachable by the agent. Walled-off systems make the rest of the stack inert; inaccessible data renders an otherwise-ready estate unusable.

Each dimension fails in its own way when AI workloads land, and the model handles none of these problems — the data handles all of them. The NIST AI Risk Management Framework's "map" and "measure" functions, and the data-quality assumptions baked into OMB's AI governance guidance, both presuppose that this work has been done.[12] In most federal data estates, it has not.

> *High scores on four of five dimensions do not average. A profile that scores well on completeness, accuracy, and accessibility but poorly on classification and lineage is not "60% ready." It is unready against the dimensions that score low.*

SECTION 02 · THE INSTRUMENT

The five-dimension readiness scorecard 02

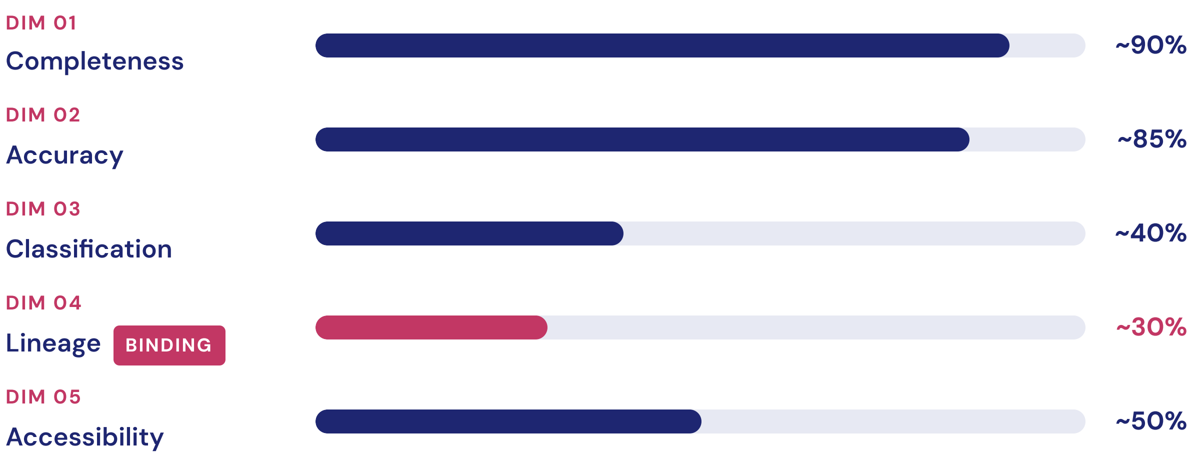
02 The five-dimension readiness scorecard

The scorecard below is the core instrument of the framework. It scores each of the five dimensions independently against the intended workload, then computes a composite — but the composite is the **floor** of the five scores, not the mean. This is the single most important and most counter-intuitive feature of the model.

FIGURE 01 · FEDERAL-ESTATE READINESS BY DIMENSION

Federal data is AI-ready only when all five dimensions hold at once.

High scores on four of five do not average. The weakest dimension is the binding constraint that decides whether the pilot ships.



Composite = the floor of the five, not the mean. A profile averaging ~59% ships at the level of its weakest input — the agent fails on lineage, not on its aggregate score. **~30%**

FCI Advisory framework, derived from federal data-quality engagement observation. Dimension levels are illustrative of the directional pattern across the federal estate, not a point-precise survey.[4]

THE FIVE-DIMENSION AI-READINESS SCORECARD

Dimension	What it measures	How it fails under AI workloads	Illustrative federal-estate level[4]
01 · Completeness	No missing fields or records across the corpus the agent will retrieve from.	Retrieval cannot fill gaps the source lacks; the model invents the missing content.	~90%
02 · Accuracy	Values reflect operational reality at the time of retrieval.	Accurate-sounding wrong answers the agent cannot self-flag.	~85%

Dimension	What it measures	How it fails under AI workloads	Illustrative federal-estate level[4]
03 · Classification	Labels and metadata that let the agent target the right content.	Unclassified content is un-queryable; retrieval returns noise.	**~40%**
04 · Lineage	Provenance is traceable from output back to source.	The audit trail is unverifiable; outputs are indefensible.	**~30%**
05 · Accessibility	The data is reachable by the agent at runtime.	Walled-off systems make the rest of the stack inert.	**~50%**
Composite	**Composite = the floor of the five, not the mean.** A profile averaging ~59% ships against retrieval workloads at the level of its weakest input. The agent fails on lineage, not on its aggregate score.		**~30%**

FCI Advisory framework, derived from federal data-quality engagement observation. Dimension levels are illustrative of the directional pattern across the federal estate, not a point-precise survey.[4]

Why the composite is the floor, not the average

A federal data estate at 90% completeness, 85% accuracy, 40% classification, 30% lineage, and 50% accessibility is *unusable* for federal AI retrieval — even though three of the five scores look fine in isolation. AI requires the weakest dimension to be strong enough; the weakest dimension is what determines whether the pilot ships. Procurement evaluations that average scores across dimensions are measuring the wrong number, and they systematically overstate readiness.

HOW TO USE THE SCORECARD

The scorecard is scored per workload, not per agency. The same estate can be AI-ready for one workload and unready for another, because the binding constraint depends on which data the workload retrieves and how it is classified. A practical readiness assessment scores each candidate workload against all five dimensions, identifies the binding constraint, and scopes remediation against that constraint first — not against whichever dimension is easiest to improve.

02·5 The five dimensions in depth

Each dimension has decades of definition behind it in records-management and database practice, and each fails in a characteristic way when an AI retrieval workload lands on it. The sections below treat each dimension as an engineering problem: what it actually requires, how it degrades, why federal estates score where they do, and what remediation looks like.

01 · Completeness

Completeness asks a deceptively simple question: does the corpus the agent retrieves from contain every record and every field the workload will need? In a human workflow, a missing field is a prompt to go find it — an analyst notices the gap and fills it. An AI retrieval system has no such instinct. When the source is incomplete, the model does not return "unknown"; it returns a fluent, confident answer assembled from whatever adjacent context it found, and that answer is wrong in a way that is hard to detect. Federal estates tend to score relatively high on completeness — records that exist are usually whole — but the failure mode is the long tail of records that were never digitized, never migrated off a decommissioned system, or never captured at all. Completeness remediation is mostly discovery and backfill: finding the gaps the source does not advertise.

02 · Accuracy

Accuracy asks whether the values in the record reflect operational reality at the moment of retrieval. This is the dimension most often confused with completeness and most often overstated. A record can be complete and still wrong — a status field that was never updated, an address that changed three years ago, a determination that was superseded. AI workloads are especially dangerous against inaccurate data because they launder it: a wrong input produces an accurate-sounding wrong answer with no visible seam between the two. Federal estates score moderately well here, but the score is workload-specific — data accurate enough for periodic reporting can be inaccurate enough to mislead a real-time agent. Accuracy remediation is reconciliation work: comparing the record of fact against the source of fact and resolving the deltas.

03 · Classification

Classification is where federal estates begin to fail. It is the metadata and labeling that lets an agent target the right content instead of retrieving everything and hoping. Without deliberate classification, a retrieval query returns noise — a mixture of relevant and irrelevant records the model cannot distinguish — and retrieval quality collapses. This is the dimension most often skipped during the original system build, because the operational workflows that created the data did not need it: a human knew which folder to open. Federal estates score low here because classification was nobody's job for two decades. It is also one of the two dimensions that most often determines whether a pilot ships. Classification remediation is the most labor-intensive of the five, and the one most amenable to AI-assisted acceleration — though that introduces its own records questions.

04 · Lineage

Lineage is provenance: the ability to trace an output back through the agent's retrieval to the specific source record, and to prove that chain to an auditor months later. It is the dimension that federal environments score lowest on and the one that matters most for defensibility. In a federal context, an action that cannot be traced is an action that cannot be defended — and an agent whose audit trail is unverifiable is, in practice, an unauthorized agent. Lineage is rarely retrofittable cheaply; it has to be instrumented at the point of data creation and carried forward. This is why lineage and classification together are the usual binding constraints, and why a remediation program that ignores them in favor of easier completeness wins is optimizing the wrong dimension.

05 · Accessibility

Accessibility asks whether the agent can actually reach the data at runtime. A perfectly complete, accurate, classified, lineage-tracked record set behind an air-gapped boundary, a deprecated API, or an identity wall the agent cannot traverse is inert — it contributes nothing to the workload. Federal estates score in the middle here, and the score is bimodal: data in modern, API-addressable systems is highly accessible, while data in mainframe-backed or heavily firewalled systems is effectively unreachable without an integration layer. Accessibility remediation overlaps with the integration-layer work covered in a companion FCI report; the readiness assessment flags it, but closing it is often a middleware problem as much as a data problem.

Cross-dimension pattern. Completeness and accuracy are usually the strongest dimensions because the original workflows needed them. Classification and lineage are usually the weakest because the original workflows did not. Accessibility sits in between and depends on the integration layer. This pattern is consistent enough across the federal estate that, absent an assessment, classification or lineage is the most likely binding constraint — and the right place to start a remediation backlog.

SECTION 03 · THE REALITY

Where federal data estates actually sit

03

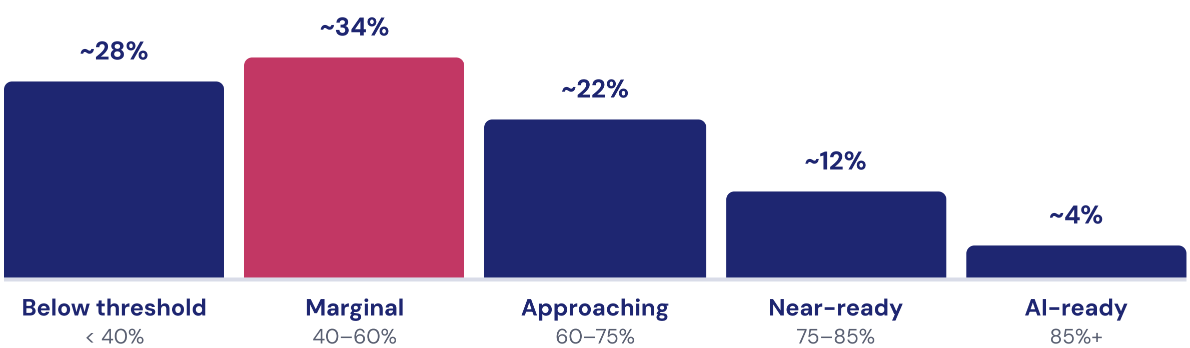
03 Where federal data estates actually sit

The federal data estate is not where the policy framework assumes it is. Across federal agencies with AI deployments in market or in pilot, the distribution of overall data readiness skews heavily toward the lower end of the maturity range. The shape of the distribution matters more than any single percentage.

FIGURE 02 · FEDERAL AI-READINESS DISTRIBUTION

The bulk of the estate sits in the lower-middle bands. Almost nothing is fully ready.

The policy assumes the right side of this chart. The operational reality is mostly the left.



Distribution of federal agencies with AI workloads in market or pilot, scored against composite readiness (weakest of five). Bands reflect FCI's engagement observation; the directional shape — not the precise percentages — is the point.[4]

ILLUSTRATIVE DISTRIBUTION OF FEDERAL AI-READINESS (COMPOSITE = WEAKEST OF FIVE)

Readiness band	Composite score	Share[4]	Operational meaning
Below threshold	< 40%	~28%	Cannot reliably pass an AI retrieval workload on any dimension.
Marginal	40–60%	~34%	Enough discipline to catalog the data; not enough to pass retrieval.

Readiness band	Composite score	Share[4]	Operational meaning
Approaching	60–75%	**~22%**	One or two dimensions still bind; targeted remediation in reach.
Near-ready	75–85%	**~12%**	Workload-specific gaps; deployable with scoped remediation.
Operational AI-ready	85%+	**~4%**	All five dimensions pass against retrieval workloads.

Distribution of federal agencies with AI workloads in market or pilot, scored against composite readiness. Bands reflect FCI's engagement observation; the directional shape — not the precise percentages — is the point.[4]

The median federal agency with an active AI deployment sits in the **marginal** band. A substantial majority sit below the "approaching" threshold on composite readiness. A small minority sit at "near-ready" or above. Almost none are fully AI-ready by the operational definition.

This is not because federal data is uniquely bad. It is because federal data was built over the past two decades against operational workflows that did not require AI-ready quality. Case management, FOIA processing, contract administration, scientific data submission — all of these tolerated data-quality issues that an automated retrieval system cannot. The data is fit for its original purpose. It is not fit for the new purpose AI workloads are creating, and nobody planned the transition. The Federal Data Strategy and the agency-level data inventories required under the Evidence Act began the cataloging work; they did not measure the five dimensions against AI workloads.[3]

Three agency archetypes

Within the distribution, three archetypes recur often enough to be worth naming. They are illustrative composites — not specific agencies — but each maps to a familiar federal pattern and to a different binding constraint.

THREE RECURRING FEDERAL DATA-ESTATE ARCHETYPES (ILLUSTRATIVE)

Archetype	Profile	Typical binding constraint	Right first move
The well-cataloged silo	Mature inventory and metadata standards, but data trapped in systems the agent cannot reach. Strong on classification, weak on accessibility.	Accessibility	Integration-layer work to expose governed data to the agent.
The clean-but-opaque estate	High-quality operational data that was never labeled for retrieval and carries no provenance. Strong on completeness and accuracy, weak on classification and lineage.	Classification, then lineage	Targeted classification of the workload corpus; instrument lineage at creation.
The fragmented landscape	Data scattered across shared drives, point systems, and email archives with no consistent inventory. Weak across all five.	Completeness, then classification	Consolidation and inventory before any AI workload is scoped.

FCI Advisory framework. Archetypes are illustrative composites used to make the binding-constraint logic concrete; real agencies sit on a continuum.[4]

The archetypes are useful because they predict the binding constraint before an assessment is run, and the binding constraint predicts where the budget should go first. An agency that recognizes itself as a "clean-but-opaque estate" already knows that a completeness-led remediation program would spend its budget on the wrong dimension — its problem is classification and lineage, not missing records.

03·5 Running the assessment

A five-dimension readiness assessment is not a months-long data audit of the entire estate. It is a focused, workload-anchored exercise that produces a defensible composite score and a ranked remediation backlog. The discipline is in scoping it tightly to the workloads the agency actually intends to run, rather than boiling the ocean of the whole data estate.

WHAT A READINESS ASSESSMENT PRODUCES

- **A per-workload scorecard.** Each candidate AI workload scored on all five dimensions against the specific corpus it will retrieve from — with the composite reported as the floor, not the average.
- **The binding constraint, named.** For each workload, the single dimension that determines whether it can ship. This is the most actionable output and the one most often skipped by conventional data-quality audits.
- **A ranked remediation backlog.** Remediation items ordered by how much they move the composite per dollar — which means weakest-dimension-first, not easiest-first.
- **A go / no-go threshold per workload.** An explicit statement of the composite score the workload needs to reach before the AI build should begin.

COMMON ASSESSMENT MISTAKES

- **Averaging the dimensions.** The single most common error. Averaging produces a number that looks reassuring and predicts nothing, because the workload fails at the floor.
- **Scoring the agency instead of the workload.** Readiness is workload-specific. An estate can be ready for one workload and unready for another; an agency-level score hides which is which.
- **Treating the assessment as a one-time event.** Data quality decays. An assessment is a snapshot; without standing governance, the score the assessment produced is stale within a year.
- **Deferring the assessment until after model selection.** By then the program is committed and the binding constraint becomes a delivery risk rather than a planning

input.

> *A readiness assessment that does not end with a named binding constraint and a go / no-go threshold has not done its job. The number that matters is not the average — it is the floor, and the date the floor clears.*

04 The cost of the gap: paid upfront vs. paid in remediation

Federal modernization budgets typically fund infrastructure first, applications second, data quality third. The sequence is almost always wrong for AI. The components of an AI program are the same regardless of order; the order is the only variable, and it is the variable that decides the cost.

FIGURE 03 · SAME COMPONENTS, DIFFERENT ORDER

The order is the variable. The right sequence costs less and ships sooner.

Identical activities, inverted. The deferred sequence pays in production; the upfront sequence pays as scoped project work.

STANDARD ORDER — WHAT MOST AGENCIES DO

STEP 01 Build data lake → STEP 02 Deploy AI workload → STEP 03 Discover quality issues → STEP 04 Reactive remediation

INVERTED ORDER — WHAT WORKS

STEP 01 Readiness assessment → STEP 02 Targeted remediation → STEP 03 Build on clean data → STEP 04 Deploy AI workload

Two sequences of the same federal AI program activities. The components are identical; the order is the only variable. Lifetime cost differences favor the inverted order by a meaningful margin across the FCI engagement base.[4]

TWO SEQUENCES, SAME COMPONENTS — THE ORDER IS THE COST DRIVER

Standard order — what most agencies do	Inverted order — what works

Standard order — what most agencies do	Inverted order — what works
1. Build the data lake / platform. 2. Deploy the AI workload. 3. Discover data-quality issues in production. 4. Fund reactive remediation.	1. Readiness assessment (five-dimension). 2. Targeted remediation of the binding constraint. 3. Build on clean data. 4. Deploy the AI workload.
Cost shape: lights-on remediation, accumulated audit obligations, pilot delivery slip. Higher total program cost, lower pilot success rate.	**Cost shape:** scoped remediation as project work, clean handoff to the AI program, predictable timeline. Lower total program cost.

FCI Advisory framework, derived from federal AI program economics observation. Lifetime cost differences favor the inverted order by a meaningful margin across the FCI engagement base.[4]

The standard sequence costs meaningfully more over the lifetime of the program, and the reasons are economic, not technical. Remediation done while AI workloads are in production has to be done with the lights on, with active users, with audit obligations already accumulated. The same remediation done before deployment is straightforward project work with clear scope and a defined end.

WHY THE WRONG ORDER IS THE POPULAR ORDER

Federal procurement economics push agencies toward the standard sequence anyway. Infrastructure spend is visible, easy to procure, and politically defensible. AI deployment spend has executive sponsorship and external visibility. Data-quality spend sits between them — necessary, invisible, hard to scope, and structurally underfunded. The wrong sequence is the default because the right sequence has no natural funding champion. The agencies that invert the sequence pay less in total, ship pilots that work, and avoid the remediation cycle that catches the agencies that did not.

> *Data-quality work is hard to scope, hard to demonstrate value for, and easy to defer when budget pressure hits. That is exactly why it is the line item most likely to be cut — and the one whose absence stalls the pilot.*

Why remediation costs more after deployment than before

The cost asymmetry between the two sequences is not a matter of a fixed amount of work being done at a slightly worse time. The same nominal remediation is structurally more expensive after a deployment than before it, for reasons that compound. Naming them makes the economics legible to a budget owner who has to defend the upfront spend.

WHAT MAKES POST-DEPLOYMENT REMEDIATION MORE EXPENSIVE

Cost driver	Before deployment	After a stalled deployment
Operating conditions	Scoped project work with a defined start and end.	Done "with the lights on" — live users, production load, no clean window.
Audit exposure	No accumulated obligations against unready data.	Audit findings and compliance obligations have already accrued.
Schedule pressure	Predictable timeline; remediation gates the build.	Under deadline pressure to un-stall a program already in the open.
Technical debt	Clean data handed to the AI build.	Workarounds built around the gap now have to be unwound first.
Stakeholder cost	An expected planning step.	A visible failure to explain and recover from.

FCI Advisory framework, derived from federal AI program cost observation. The total-cost gap favors the upfront sequence by a meaningful margin across the FCI engagement base.[4]

None of these drivers is hypothetical, and none of them appears in a model-benchmark comparison. They are the reason the inverted sequence is cheaper in total even though it spends earlier — the upfront sequence pays a bounded, predictable cost; the deferred sequence pays an unbounded one under worse conditions. The economics are the same shape as the records-modernization economics covered in a companion FCI report: bounded upfront, compounding when deferred.

05 The maturity model

Federal data readiness for AI is not binary. It moves through a defined sequence of four stages, each with characteristic capabilities and characteristic gaps. The model is descriptive, not prescriptive — most agencies do not skip stages, and the transition between adjacent stages takes between roughly 6 and 24 months depending on funded scope.

THE FOUR-STAGE FEDERAL DATA-READINESS MATURITY MODEL

Stage	Characteristic capability and gap	Illustrative share[4]	Composite
01 · Ad hoc	Data in operational silos. No cross-agency inventory. Quality varies by team and is undocumented. AI workloads cannot reliably retrieve.	**~38%**	< 40%
02 · Cataloged	Data inventory exists; some metadata standards. Quality measured opportunistically on one or two dimensions. Lineage and classification mostly absent.	**~44%**	40–60%
03 · Governed	Five-dimension quality measurement in place. Lineage tracked across systems. Classification deliberate. Remediation programs funded and operational.	**~14%**	60–85%
04 · AI-ready	All five dimensions pass against retrieval workloads. Audit trail complete. Sustained governance with ongoing measurement.	**~4%**	85%+

Four-stage descriptive model. Distribution percentages reflect aggregate FCI observation; specific agency placement varies by workload and dimension. Stages are sequential.[4]

FIGURE 04 · THE MATURITY MODEL

Most agencies sit in Stages 1 and 2. Stage 2→3 is the reachable move.

Stage 2 to Stage 3 fits inside a 12–18 month remediation program — and is the transition this budget cycle should fund.

STAGE 01	STAGE 02	STAGE 03	STAGE 04
Ad hoc	**Cataloged**	**Governed**	**AI-ready**
Data in operational silos. No cross-agency inventory. Quality varies by team. AI workloads cannot reliably retrieve.	Inventory exists; some metadata standards. Quality measured on one or two dimensions. Lineage and classification mostly absent.	Five-dimension measurement in place. Lineage tracked across systems. Classification deliberate. Remediation funded and operational.	All five dimensions pass against retrieval workloads. Audit trail complete. Sustained governance with ongoing measurement.
~38% federal agencies	**~44%** federal agencies	**~14%** federal agencies	**~4%** federal agencies

Four-stage descriptive model of federal data readiness for AI. Distribution percentages reflect aggregate FCI observation; specific agency placement varies by workload and dimension.[4]

Most federal agencies sit in Stage 1 or Stage 2 today. Stage 3 is reachable inside a 12–18 month remediation program with focused effort and budget. Stage 4 — full operational AI-readiness — takes longer and requires sustained governance investment, not a one-time cleanup project. Each transition has predictable cost and predictable duration; the variability is mostly in how much the agency chooses to fund.

> *The agencies deploying federal AI successfully over the next three years are the agencies that move from Stage 2 to Stage 3 in this budget cycle. The agencies that defer that transition will discover the gap when their workloads hit production — and remediate under worse economics.*

06 A staged remediation sequence

The remediation sequence below is the inverted order made concrete. It is scoped to move a typical Stage-2 agency to Stage 3 — the realistic, fundable transition for most of the federal estate — and to do it before, not after, an AI workload lands.

STEP 01 · WEEKS 0–6

Workload-anchored readiness assessment

Score each candidate AI workload against all five dimensions. Identify the binding constraint per workload. Output: a per-workload scorecard and a ranked remediation backlog. Do not start with a platform; start with the workload and the constraint.

STEP 02 · MONTHS 2–6

Targeted remediation of the binding constraint

Remediate the weakest dimension first, because the composite is the floor. For most federal estates that means classification and lineage — the two dimensions that score lowest and are most often skipped. Completeness and accuracy fixes follow only where they bind a specific workload.

STEP 03 · MONTHS 4–9

Stand up lineage and classification as standing capabilities

Instrument provenance and metadata at the point of data creation, not retroactively. This is the step that distinguishes a one-time cleanup from a governed estate — and the step that keeps the estate ready as new data arrives.

STEP 04 · MONTHS 6–12

Build the AI workload on clean, governed data

Only now does the AI build begin, on an estate whose binding constraint has been removed. The pilot is scoped to data that has already passed the five-dimension test for that workload.

STEP 05 · ONGOING

Sustained governance, not a closed project

Data quality decays without governance. Stand up the standing function that maintains the five dimensions over time. Agencies that fund a one-time remediation without the governance function see the same gap reappear within two years.

Sequencing principle. Remediate the binding constraint, not the easiest dimension. The composite is the floor of the five — so a dollar spent on the weakest dimension moves readiness; a dollar spent on an already-strong dimension does not. This is the opposite of how most data-quality programs allocate effort.

The governance operating model

Stage 3 is defined less by a finished cleanup than by a standing function that keeps the five dimensions from decaying. The most common reason a remediation program fails to hold is that it was scoped as a project with an end date rather than as a capability with an owner. The operating model below is the minimum standing structure that keeps a governed estate governed.

MINIMUM GOVERNANCE OPERATING MODEL FOR A STAGE-3 ESTATE

Function	Responsibility	Why it has to stand, not stop
Data-quality measurement	Re-score the five dimensions on a recurring cadence against the active workloads.	Quality decays as new data arrives and systems change; an annual snapshot misses the drift.
Classification at creation	Label and tag new data when it is created, not retroactively.	Retroactive classification is the most expensive remediation; doing it once does not stop the backlog from rebuilding.
Lineage instrumentation	Maintain provenance capture across the systems the agent retrieves from.	Lineage cannot be reconstructed after the fact; it has to be carried forward continuously.
Workload intake	Score each new candidate AI workload before it is funded.	Every new workload has its own binding constraint; readiness is not a one-time clearance.
Records-officer alignment	Coordinate with the records function on classification, retention, and AI-generated outputs.	AI outputs are themselves federal records; data governance and records governance converge.

FCI Advisory framework, derived from federal data-governance engagement observation.[4]

The last row matters more than it appears. As AI workloads land, the data-readiness function and the federal records function stop being separate concerns — the agent's outputs become records, and the same classification and lineage discipline that makes data AI-ready also makes AI outputs records-ready. Agencies that run these as one governance conversation, rather than two, avoid building a readiness program that quietly creates an unmanaged records problem. FCI treats the two as a single architecture; this report's companion on the federal records mandate develops that convergence in full.

What this rules in and out

Four strategic conditions reshape what federal CIOs should be funding through FY26 and into the next budget cycle:

- **Data quality is a procurement prerequisite, not a downstream concern.** Every federal AI procurement should include a data-readiness assessment scoped before model selection, with remediation funded as part of the program — not as a separate later project. Programs that defer the data work pay for it twice and ship later than planned.
- **The five dimensions are non-substitutable.** Data that scores high on four of five is not 80% ready; it is unready against the dimension that scores low. Procurement evaluations that average scores across dimensions are measuring the wrong number.
- **Stage 2 to Stage 3 is the realistic next move for most agencies.** Trying to leapfrog from Stage 1 or 2 directly to Stage 4 overcommits budget on capabilities the agency cannot yet operate. The deliberate path moves through stages; budget cycles should align to stage transitions.
- **Sustained governance matters more than one-time cleanup.** Data quality decays. Agencies that fund a cleanup without standing up the governance function will see the same gap reappear within two years. The cleanup is the start of the work, not the end.

> *The decision for federal CIOs is whether to fund the data-readiness work as a deliberate prerequisite, paid upfront with clear scope, or as a forced remediation after the pilot has failed to ship. The cost is similar either way. The timing changes whether the program delivers.*[4]

Put this thinking to work.

FCI Advisory helps federal agencies turn analysis like this into delivered outcomes — from five-dimension readiness assessments to funded, sequenced remediation programs. McLean, Virginia.

Endnotes & sources

1. OMB Memorandum M-24-10, "Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence," March 28, 2024.
2. NIST AI Risk Management Framework (AI RMF 1.0), National Institute of Standards and Technology — specifically the "map" and "measure" function categories addressing data quality, lineage, and representativeness.
3. The Federal Data Strategy and the Federal Data Strategy Action Plans published by OMB; agency-level data inventories under the Foundations for Evidence-Based Policymaking Act (the "Evidence Act").
4. Aggregate observations cited in this report are drawn from FCI Advisory's engagement base across federal data-quality and AI-readiness programs, FY24-Q4 through FY26-Q1. Distribution figures, dimension levels, and stage percentages are illustrative of the directional pattern, not point-precise extracts.