



INTELLIGENT ARCHIVING

THE INTELLIGENT ARCHIVING FRAMEWORK

A practical framework for managing long-term Microsoft 365 content that is trusted, compliant, and AI-ready.



INTRODUCTION

AI is transforming how organizations discover, enrich, and analyze information. But the rush to leverage AI-powered insights has created a dangerous blind spot: the assumption that AI can work effectively on information that has not been properly managed. It cannot.

Before information can be surfaced by Microsoft Copilot, enriched by machine learning, or analyzed by large language models, it must be governed: controlled, compliant, readable, and defensible. Unmanaged cloud content, legacy file formats, and unclear retention rules do not become less risky when AI is introduced. The risk is that AI amplifies whatever it finds, including errors, outdated records, and compliance gaps. According to AIIM's 2025 Industry Watch research, 77% of organizations rate their data quality as average, poor, or very poor for AI readiness, meaning most organizations are exposing AI to content that is already untrustworthy. This makes governance gaps that were once merely inefficient, actively dangerous.

This framework outlines a practical five-step approach for getting your information in order and a Governance Readiness Scorecard to help you assess where your organization stands today.



FIVE STEPS TO AI-READY INFORMATION

Getting your information ready for AI does not require a single massive transformation. It can be approached as a structured, incremental process. The following five steps provide a practical roadmap for getting your information in order.

STEP 1: CLASSIFY AND APPLY RETENTION

Establish clear classification and retention rules across your Microsoft 365 environment. Leverage Microsoft Purview retention labels and policies to identify what content exists, how long it must be retained, and what disposition action applies when retention expires.

Key actions:



This step requires deliberate collaboration: records and compliance professionals define the policy, IT teams make it enforceable in Microsoft 365.

AIIM's 2025 Industry Watch research found that governance responsibility is typically fragmented across eight or more organizational functions, making clear role definition a prerequisite, not an afterthought.

Organizations that complete this step typically discover significant volumes of redundant, obsolete, or trivial (ROT) content that wastes storage costs and creates compliance exposure if surfaced by AI.

STEP 2: ARCHIVE & PRESERVE LONG-TERM CONTENT

Once content is classified, separate long-term and permanent records from active operational content. Transfer them from active SharePoint libraries into a dedicated preservation archive purpose-built for ensuring readability, integrity, and compliance over decades.

This step simultaneously:



"Manual processes are very labor-intensive. Automation maintains the integrity of our records and provides an efficient and easy way for ensuring permanent archival records in our statewide Microsoft 365 tenancy are preserved and accessible in-place regardless of legal custodianship."

State Archivist and Chief Records Officer, US state government agency.



Automation is essential. By leveraging Microsoft Power Automate and Purview labels, organizations can create automated workflows that transfer content to the preservation archive based on retention rules, minimizing manual intervention and human error.

STEP 3: CONSOLIDATE SILOED LONG-TERM CONTENT

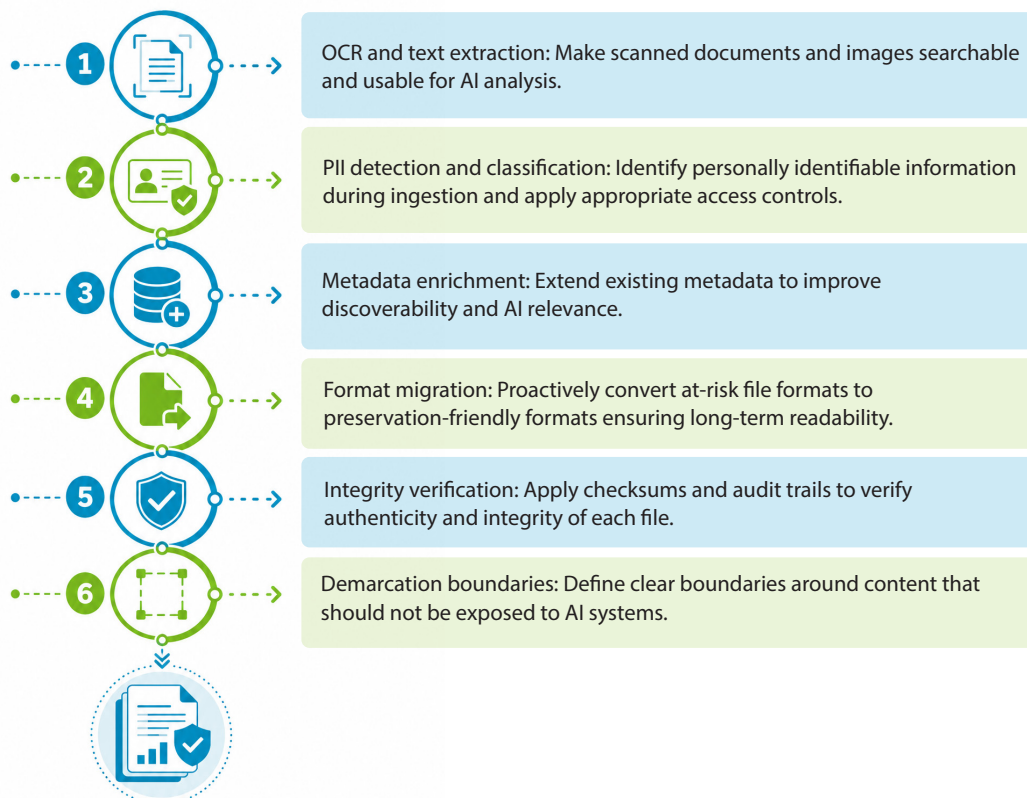
Many organizations have critical long-term content trapped in legacy systems, including retired ECM platforms, file shares, email archives, and proprietary databases. According to AIIM's 2025 Industry Watch research, 76% of organizations struggle with siloed or fragmented data. This content is often invisible to modern search tools, inaccessible to AI, and at risk of format obsolescence or data loss.

A governance-first approach includes consolidating this legacy content into a dedicated preservation archive embedded within the Microsoft 365 ecosystem. This makes SharePoint a discovery hub for both current Microsoft-native content and historical information from non-Microsoft systems, potentially dating back 10, 15, or 20 years.

Double down on your Microsoft 365 investment by making SharePoint not just a collaboration platform but a unified discovery hub for all organizational information, current and historical.

STEP 4: ENRICH AND PROTECT CONTENT QUALITY

With content classified, archived, and consolidated, enrich its quality and ensure it remains trustworthy over time:



STEP 5: ENABLE SAFE AI DISCOVERY

Only after the preceding steps are complete is your information truly ready for AI. With governed, preserved, and enriched content, you can now safely enable AI-powered capabilities:

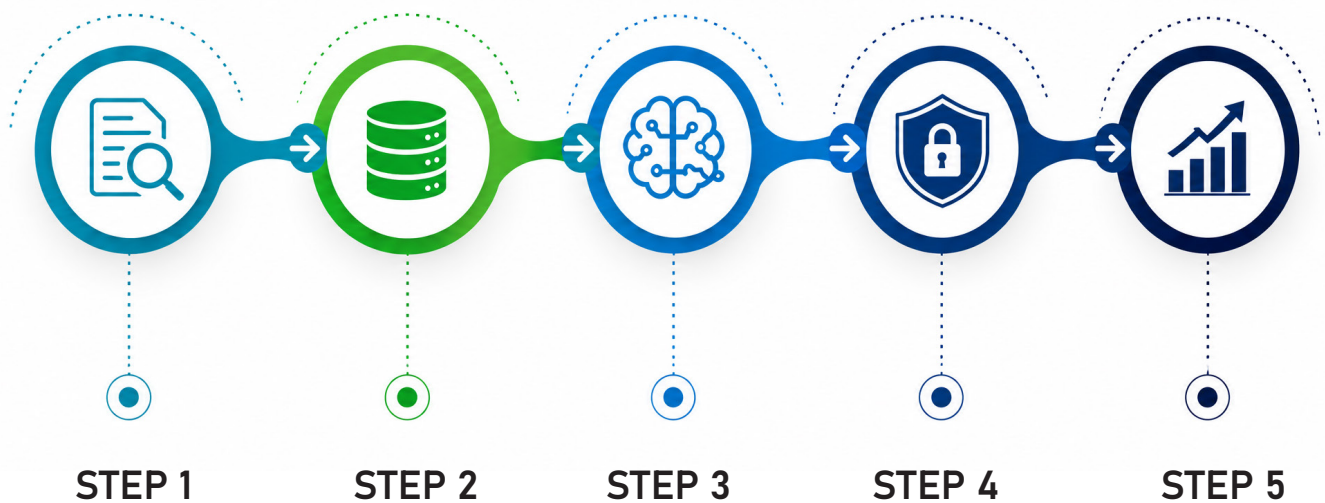
- ◆ Microsoft Copilot surfaces archived content through SharePoint, grounded in trustworthy information
- ◆ AI-powered tools enhance metadata, improve classification accuracy, and generate summaries
- ◆ Large language models produce reliable outputs from high-quality, trusted data
- ◆ Content discovery is faster and more accurate on a clean, well-governed foundation

The critical distinction: AI is applied as an enhancement on top of a solid governance foundation, not deployed as a shortcut around governance. This ensures that AI-generated insights are as reliable as the information they are built on.



PART 2

GOVERNANCE READINESS CHECKLIST



GOVERNANCE READINESS CHECKLIST

This checklist helps you evaluate your organization's readiness to manage long-term digital content for both compliance and AI. Use it to benchmark your current state, identify gaps, build your roadmap, and track progress over time.

Rate each statement against your current state: 0 = not in place, 1 = partially in place, 2 = fully in place. Total each section, then add up your overall score.

	0 – Not in Place	1 – Partial	2 – In Place
STEP 1: CLASSIFY & APPLY RETENTION			
Retention schedules are actively enforced across all M365 content types, not just documented – with evidence of consistent application and regular review.			
Long-term and permanent records are separated from active content into a dedicated preservation archive.			
Archived content is protected against format obsolescence so it remains readable over decades.			
Section Score: _____ / 6			

	0 – Not in Place	1 – Partial	2 – In Place
STEP 2: ARCHIVE & PRESERVE LONG-TERM CONTENT			
Long-term and permanent records are separated from active content into a dedicated preservation archive.			
Long-term and permanent records are separated from active content into a dedicated preservation archive.			
Automated workflows transfer records to the archive based on retention rules, with full audit trails.			
Section Score: _____ / 6			

	0 – Not in Place	1 – Partial	2 – In Place
STEP 3: CONSOLIDATE SILOED LONG-TERM CONTENT			
Legacy systems (retired ECM platforms, file shares, email archives) holding long-term content have been inventoried.			
Historical content from non-Microsoft systems is consolidated and discoverable through SharePoint.			
Section Score: _____ / 4			

	0 – Not in Place	1 – Partial	2 – In Place
STEP 4: ENRICH & PROTECT CONTENT QUALITY			
Scanned documents are OCR-processed and searchable so they are usable by both people and AI.			
Checksums, file identification, and audit trails verify the authenticity and integrity of preserved content.			
File-level metadata (creator, classification, retention category) is accurate, consistently applied, and verifiable – not just present at the folder or library level.			
Clear boundaries define which content must not be exposed to AI tools due to PII, legal privilege, or sensitivity – and those boundaries are technically enforced, not just policy statements.			
Section Score: _____ / 8			

	0 – Not in Place	1 – Partial	2 – In Place
STEP 5: ENABLE SAFE AI DISCOVERY			
AI tools such as Microsoft Copilot access only governed, preserved, and quality-verified content.			
A documented policy governs how AI interacts with long-term records, including chain of custody requirements for content surfaced or acted upon by AI, and is reviewed as AI capabilities evolve.			
Section Score: _____ / 4			
Total Score: _____ / 28			

YOUR SCORE

SCORE	READINESS	WHAT IT MEANS
0 – 8	Foundation	Significant governance gaps. AI tools will amplify risk if deployed now. Prioritize Steps 1 and 2.
9 – 16	Developing	Governance underway but inconsistent. Complete classification, archiving, and legacy consolidation before expanding AI use.
17 – 22	Established	Core governance in place. Focus on content enrichment, format preservation, and defining AI boundaries.
23 – 28	Ready	Strong governance foundation. Safe to enable AI discovery on trusted, preserved content. Monitor and optimize.

YOUR ROADMAP

The steps are sequential. Gaps in earlier steps undermine everything that follows. Governance first, intelligence second. The order matters.

WHERE YOU SCORED LOW	PRIORITY	WHAT TO DO NEXT
Step 1 scored low	Start here	Audit content. Define retention schedules. Apply retention labels and enforce policies. Remove ROT before AI surfaces it.
Step 2 scored low	High	Move long-term records to a preservation archive. Automate transfers. Ensure format readability over decades.
Step 3 scored low	High	Inventory legacy systems. Consolidate into a unified archive discoverable through SharePoint.
Step 4 scored low	Medium	OCR scanned content. Verify integrity with checksums and audit trails. Define AI access boundaries.
Step 5 scored low	Do last	Enable AI only on governed, preserved content. Document an AI governance policy. Review it annually.

Revisit this checklist quarterly to track progress and adjust priorities as your governance maturity and AI capabilities evolve.



CONCLUSION

The AI era has created unprecedented opportunities for organizations to extract value from their information. But it has also raised the stakes for information governance. Poorly managed content does not become less risky when AI is introduced; it becomes riskier.

The organizations that will succeed are those that adopt a governance-first approach: classifying their content, applying retention rules, archiving long-term records to preservation systems, consolidating legacy content, enriching quality, and only then enabling AI discovery on a trusted foundation.

Your long-term content has unprecedented strategic value. By governing it properly, you simultaneously strengthen your compliance posture and position your organization to create value through AI, safely, confidently, and with full control.

ABOUT PRESERVICA

Preservica is changing the way organizations around the world protect and future-proof long-term digital information. Preservica's unique Active Digital Preservation archiving software automatically keeps every file alive in future-friendly formats over decades to ensure high-value information can always be quickly found, trusted, and actioned for FOI, regulatory compliance, legal, brand, and AI needs.

Developed in collaboration with Microsoft, Preservica's Preserve365 simplifies compliance and AI adoption by making the archiving, digital preservation, and retrieval of long-term and permanent records part of everyday Microsoft 365 and Microsoft SharePoint workflows. This seamless, embedded approach empowers Microsoft customers to maximize existing software investments such as Purview, Power Automate, and Copilot to automate archiving, reduce cost and risk, and build a foundation of trusted AI-ready long-term content.

Learn more at [Preservica.com](https://www.preservica.com)

ABOUT AIIM

AIIM is the authority on data and information readiness, governance, and value-delivery for AI. AIIM equips information leaders with practical use cases, certification, training, and guidance to leverage information management for AI, process automation, and emerging technologies. The Certified Information Professional (CIP) credential validates the complete lifecycle expertise today's information leaders need, from data mapping and quality assurance through AI governance and strategic leadership.

Learn more at www.aiim.org