aiim
Association for Intelligent
Information Management

Microsoft 365

Unlocking the Value of Long-term Microsoft 365 Content for Compliance and AI

P Preservica

Contents

Introduction

Organizations across government and regulated industries face a dual imperative: ensuring compliance and operational efficiency today while simultaneously positioning themselves to harness the power of AI (Artificial Intelligence).

The growing volume of content and records that organizations create or migrate into collaboration platforms like Microsoft SharePoint compounds this challenge, as up to one-third of this content requires retention not just for years, but for decades, according to analyst firm IDC.

This guide explores how the proactive management of this critical long-term content can drive efficiency and cost savings today as well as become an essential and trusted data foundation for both defensive compliance needs and offensive value creation through AI.

IDC Info Snapshot 2024 reports that

One-third

of data in Microsoft 365 needs to be retained long term

What is Long-Term Content?

Long-term content refers to information that organizations must retain for extended periods, typically 7 years or more, and often permanently. According to Preservica's research, information and records managers report that 20-28% of their records have retention periods exceeding 7 years or are permanent.

The long-term content that needs to be retained will obviously differ between industries and government agencies, but can include:

- **Board, committee and strategic decision making**

- **Engineering specifications, building plans and drawings**

- **Leases, permits and contracts**

- **Intellectual property and trademarks**

- **Product packaging and datasheets**

- **Health & Safety and Environmental reports**

- **HR and Personnel files**

- **Pension and Life Insurance documentation**

- **Vital records**

According to Preservica, this category of information comprises both simple individual file formats (like PDFs) and more complex "multipart" formats such as emails with attachments, AV content, CAD, 3D files and Microsoft Teams. What distinguishes long-term content from other information categories is that it must be stored, accessed, trusted and used over many years and through multiple technology cycles and custodians.

Records managers said more than

20-28%

of their records have >7 years retention or are permanent

"Optimizing the Automated Governance Capabilities of Microsoft 365" – MER, 2022

The Strategic Value of Long-Term Content

Long-term content holds significant value for organizations in four key areas:

Regulatory Compliance:
Information is retained to comply with externally applied industry rules and legal mandates. Failure to produce trustworthy information in a timely fashion can lead to substantial fines.

Legal Protection:
Information may be needed for evidential purposes, liability protection and reputational risk management.

Operational Efficiency:
Information is retained to support ongoing operational needs such as maintenance, Health & Safety improvements or contractual negotiations

Competitive Advantage:
Long-term information, especially in the AI era, has inherent strategic value that can be leveraged to optimize processes and provide new insights
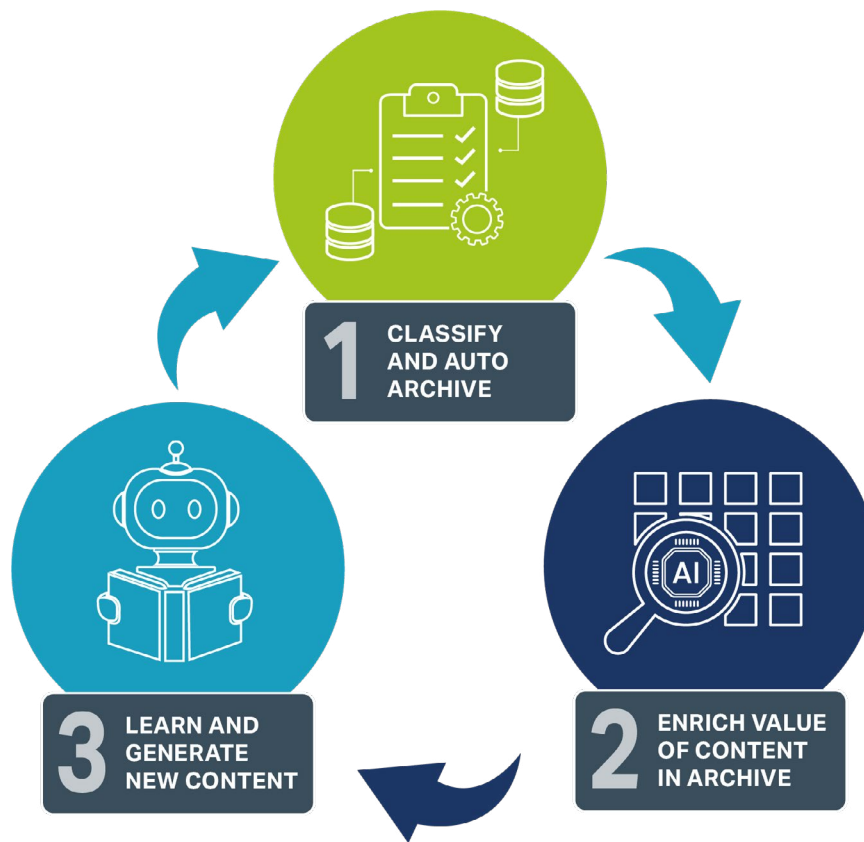
In the AI era, long-term content is no longer just a necessary compliance burden or cost. Organizations depend on the unstructured data found in long-term content to train AI models and complete AI tasks. Future-focused organizations recognize that long-term content represents a foundation of organizational knowledge that can drive innovation, enhance decision-making, and create competitive advantage.
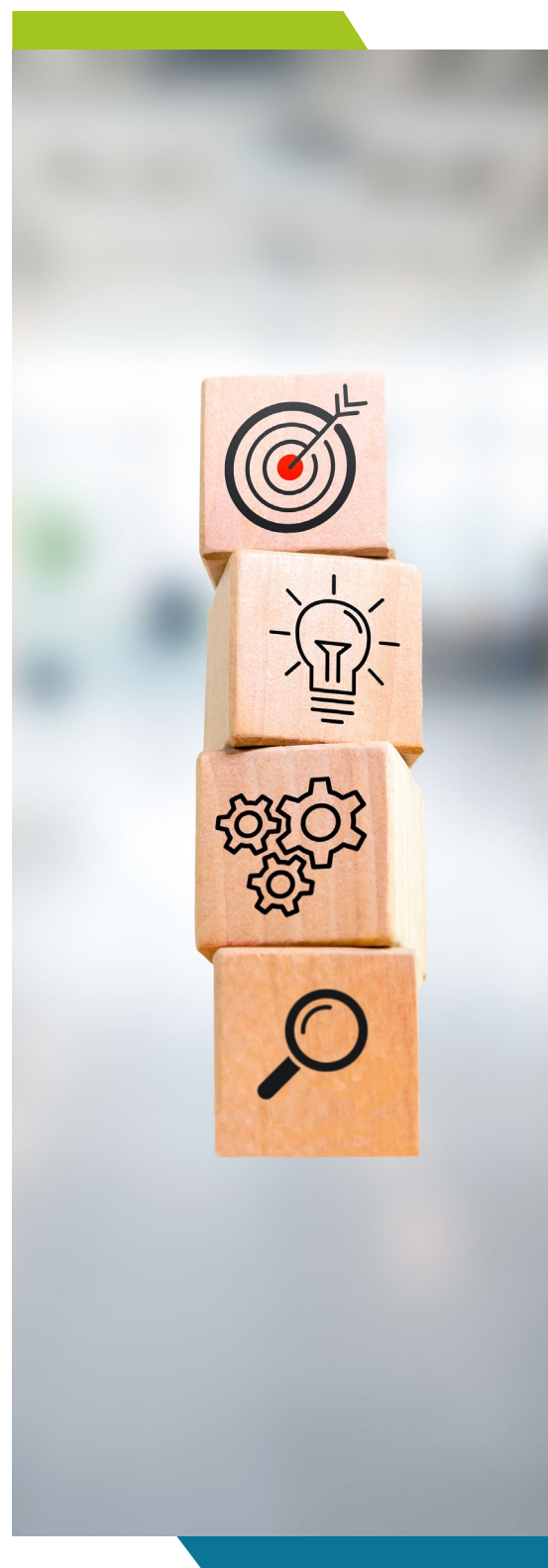
Management of Long-Term Content in Microsoft 365

With the growth of Microsoft 365 adoption and the consolidation of legacy systems into Microsoft SharePoint, a particular focus needs to be placed on how the growing volume of long-term and legacy content is managed and maintained in an efficient way within the Microsoft ecosystem.

Leveraging Microsoft labels and tools for automation and retention are essential for compliance accuracy and efficiency. Combining these with ways to automate the archiving of long-term content can reduce SharePoint costs and clutter and ensure that only the most accurate and reliable final records are transferred to the archive – building a foundation of trusted, high-quality data essential for compliance and reliable AI outcomes.

1 CLASSIFY AND AUTO ARCHIVE

2 ENRICH VALUE OF CONTENT IN ARCHIVE

3 LEARN AND GENERATE NEW CONTENT

The Importance of Quality Long-Term Content for AI Success

The quality of an organization's data is fundamental to successful AI implementation. According to AIIM's *State of the Intelligent Information Management Practice 2024 Report*, data quality remains a significant obstacle to AI adoption, with 77% of respondents rating their organizational data as either average, poor, or very poor in terms of quality and readiness for AI.

For AI to deliver accurate, reliable, and optimal outcomes - whether for value creation, learning, process optimization, or decision-making - the underlying long-term content must be properly archived and preserved to ensure it is inherently trustworthy and usable.

Data quality remains a significant obstacle to AI adoption, with

77%

of respondents rating their organizational data as either average, poor, or very poor in terms of quality and readiness for AI.

"Intelligent Information Management Practice Report" - AIIM, 2024

Common Threats to Long-Term Content Quality

Several threats can compromise the quality and usability of long-term content:

- **Format Obsolescence**: If the original software needed to read a file is not accessible, the file becomes unreadable – for example try watching the Real Media .rm files you might have migrated into SharePoint.

- **Integrity and Provenance:** Being able to maintain and demonstrate the inherent integrity of long-term content is essential to reliable AI outcomes and vital in a world of AI generated fakes and hallucinations.

- **Context Loss**: When information is composed of multiple components, such as files and metadata losing just one element can fundamentally change how the information can be used and interpreted.

- **Discoverability**: Information that cannot be efficiently found and accessed might as well not exist.

- **Permission Loss**: Information may become inaccessible to users who need it.

- **License Loss**: Even if software is still available, lost license keys render it unusable.

- **Change or deletion**: Information retained in its original system may be edited or deleted after the decision to preserve it.

- **Siloed or Fragmented Data**: AIIM research shows that 76% of organizations face this issue, which leads to inconsistencies and integration difficulties.

- **Redundant, Obsolete, or Trivial (ROT) Data**: According to AIIM, 76% of organizations struggle with duplicated or unnecessary data.
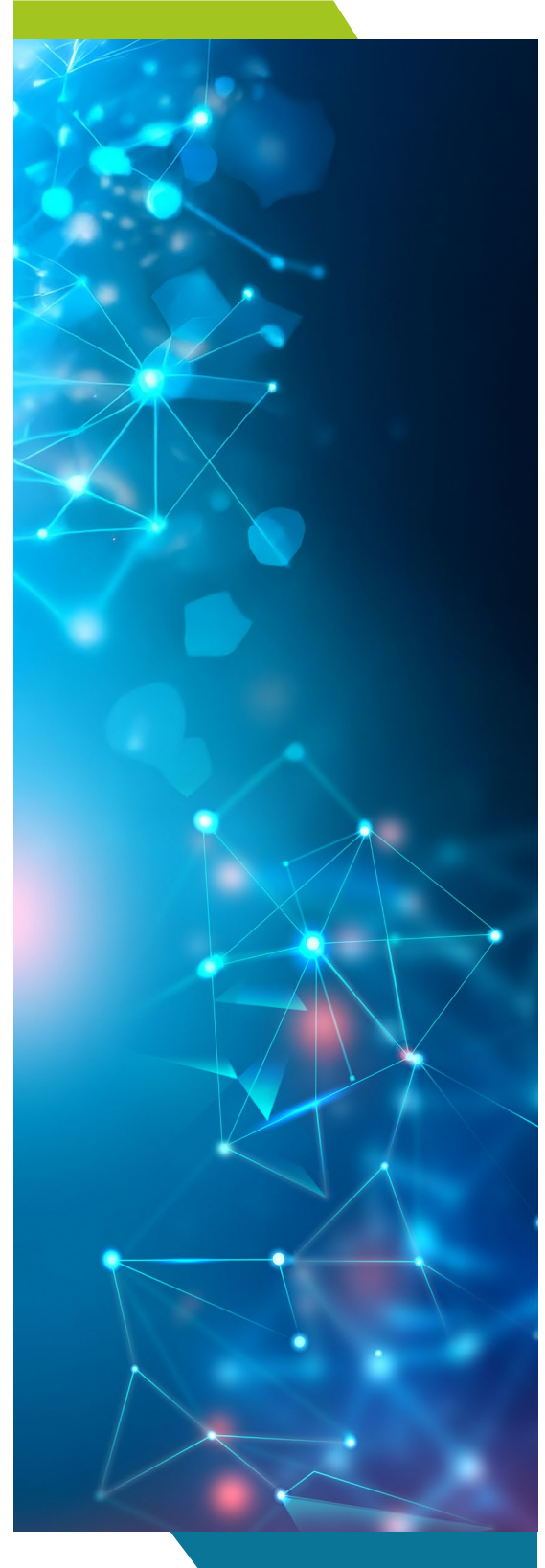
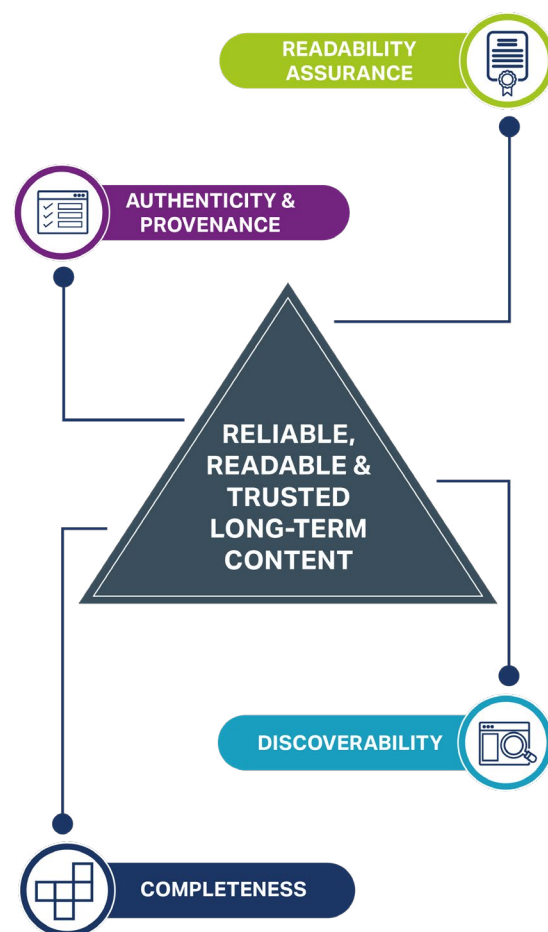# Protecting Long-Term Content quality to Ensure Ongoing Reliability

Conventional backups, archiving, and enterprise content management systems used for short and medium-term content retention fall short when it comes to ensuring information remains usable and trusted over extended timeframes of 10, 20, 30, or even 100 years. These systems were designed for operational efficiency, not long-term preservation.

Having a complementary archiving and preservation strategy that addresses the unique requirements of protecting long-term content is essential for compliance and AI use. These unique requirements include:

- **Readability Assurance**: Files must be continuously protected against format obsolescence through format migration, ensuring they can be opened and read even when original applications are no longer available.

- **Trust & Authenticity**: Each file's integrity and authenticity must be verifiable through mechanisms like checksums, file identification, and comprehensive audit trails of every action taken.

- **Discoverability**: With proper access controls, long-term content must remain searchable so the right information can be found quickly and understood in its original context.

- **Completeness**: Complex information packages with multiple files and metadata must be maintained as complete units to preserve essential context while protecting each individual file's integrity.

Long-term digital preservation combines precision strategies specifically designed to mitigate risks associated with threats to long-term content, ensuring information remains accessible, usable, and trustworthy far into the future.

While standard archiving focuses on storage efficiency, true long-term preservation addresses format obsolescence, bit-level integrity, contextual relationships, and authenticated access—transforming archiving from merely a compliance burden into a valuable strategic asset.

READABILITY ASSURANCE

AUTHENTICITY & PROVENANCE

RELIABLE, READABLE & TRUSTED LONG-TERM CONTENT

DISCOVERABILITY

COMPLETENESS

Assessment Scorecard

Introduction to Long-Term Content Maturity Assessment: Compliance & AI Readiness

Organizations must maintain digital content not only for compliance and regulatory purposes but also ensure it remains usable, trusted, and optimized for future value creation through AI initiatives. This assessment will help you evaluate your organization's maturity in these critical capabilities.

The following capabilities represent the foundation of an effective long-term Microsoft 365 content strategy, both compliance requirements and AI readiness. Each capability corresponds to a question in the assessment that follows

1 COMPREHENSIVE PRESERVATION

Beyond simple storage, implement solutions that actively prevent file format obsolescence, prove authenticity, and ensure accessibility over decades. This fundamental capability safeguards your organization's digital memory against technological change and lays the foundation for compliance and AI.

2 DATA INTEGRITY & AUTHENTICITY

Establish processes that allow each file's integrity and trustworthiness to be easily proven through checksums, audit trails, and validation methods. This capability is essential for compliance, legal defense, and trusted AI inputs.

3 CONTEXTUAL INTEGRITY

Retain the full context of your information as complete packages, preserving relationships between related files and metadata. Contextual understanding is crucial for both compliance documentation and effective AI implementation.

4 ON-DEMAND RETRIEVAL & USABILITY

Ensure archived content can be easily searched, retrieved, and used on-demand, even when original file formats become obsolete or unsupported by business and legal teams. Immediate access to trusted information provides competitive advantage.

5 AUTOMATED ARCHIVING PROCESSES

Be able to leverage existing Microsoft tools to automate records transfer to preservation systems, minimizing manual intervention and human error. Automation improves consistency while reducing operational costs.

6 CONTENT QUALITY & HYGIENE

Systematically eliminate inappropriate language, outdated information, and trivial content (ROT data) to improve search relevance and usability. Clean, relevant content is essential for both compliance and AI effectiveness.

7 METADATA & TAXONOMY

Develop well-designed metadata and taxonomy plans that make content more discoverable and can be weighted appropriately for relevance in search results and enhances both human and AI interaction.

8 TECHNOLOGY INTEGRATION

Maximize existing software and technology investments by making preservation capabilities part of existing compliance and retrieval workflows. Well-integrated systems create efficiency and automatically build out quality trusted long-term content.

Assessment Scorecard

9 DATA QUALITY FOR AI

Implement automation processes to ensure only the best quality data enters the preservation archive for (compliance) and AI use. Then address missing data, outliers, standardization issues, and potential bias. Quality data inputs lead to more reliable AI outputs.

10 DISCOVERABILITY OPTIMIZATION

Enhance content findability through AI-powered enrichment techniques and extend metadata with AI provenance information. Enhanced discoverability multiplies the value of preserved content.

11 GOVERNANCE & ACCESS CONTROL

Define clear permission boundaries between human and AI access, determining what data should not be exposed due to policy, ethics, or proprietary concerns. Appropriate governance balances innovation with risk management.

12 HUMAN OVERSIGHT & ENHANCEMENT

Implement feedback loops and human validation for sensitive classifications and content enrichment processes. Human oversight ensures AI systems remain aligned with organizational values and goals.

13 PRIVACY & SECURITY PROTECTION

Deploy PII classification and identification during content ingestion with processes to manage sensitive content appropriately. Proper privacy controls protect both individuals and organizational reputation.

14 STRATEGIC VALUE RECOGNITION

View long-term content preservation as a strategic asset rather than just a compliance requirement. Organizations that recognize the strategic value of preserved content gain competitive advantage through both risk mitigation and AI-driven innovation.

SUMMARY

By assessing your organization's capabilities in these areas, you'll gain valuable insights into your current maturity level and identify opportunities for improvement. Organizations that establish a solid foundation of trusted, usable long-term content will be better positioned to both meet compliance requirements and leverage AI for competitive advantage.

Assessment Scorecard

Long-Term Content Maturity Assessment: Compliance & AI Readiness

This assessment is designed to help your organization evaluate its current capabilities in preserving and managing long-term digital content for compliance and AI.

How to Use This Assessment:

Benchmark Your Current State:
Complete this assessment to establish your organization's baseline digital preservation and long-term content quality maturity.

Identify Gaps and Opportunities:
The questions highlight critical capabilities needed for effective long-term content management. Lower scores indicate areas requiring immediate attention.

Build Your Roadmap:
Use the scoring guide and recommendations to prioritize investments and improvements in your digital preservation strategy.

Engage Stakeholders:
Share results with IT, compliance, records management, and executive teams to create alignment on the strategic importance of digital preservation.

Track Progress:
Re-assess periodically (e.g., annually) to measure improvement and adjust your strategy as needed.

Your long-term content holds strategic value far beyond meeting compliance standards in this AI-driven era. Organizations that establish a solid foundation of trusted, usable long-term content now will be better positioned to leverage AI and create competitive advantage.

Assessment Scorecard

Instructions

Rate your organization's capabilities on a scale of 1-5 for each statement below:

1 = Strongly Disagree   2 = Disagree   3 = Neither Agree nor Disagree   4 = Agree   5 = Strongly Agree

Assessment	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
COMPREHENSIVE PRESERVATION Our organization implements preservation techniques that actively prevent file format obsolescence, prove authenticity, and ensure accessibility over decades.	1	2	3	4	5
DATA INTEGRITY & AUTHENTICITY When called upon, each file's integrity and trustworthiness can be easily proved through checksums, audit trails, or other validation methods.	1	2	3	4	5
CONTEXTUAL INTEGRITY The full context of our information is retained as a complete package, preserving relationships between related files and metadata.	1	2	3	4	5
ON-DEMAND RETRIEVAL & USABILITY Archived content can be easily searched, retrieved, and then used on-demand - even if original file formats are now obsolete or unsupported by business and legal teams.	1	2	3	4	5
AUTOMATED ARCHIVING PROCESSES We can leverage our Microsoft tools to automate records transfer to preservation systems, minimizing manual intervention and human error.	1	2	3	4	5
CONTENT QUALITY & HYGIENE We systematically eliminate inappropriate language, outdated information, and trivial content (ROT data) to improve search relevance and usability.	1	2	3	4	5
METADATA & TAXONOMY FRAMEWORKS We have well-designed metadata and taxonomy plans that make content more discoverable in search and can be weighted appropriately for relevance.	1	2	3	4	5

Assessment Scorecard

Assessment	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
TECHNOLOGY INTEGRATION We maximize our existing technology investments, including Microsoft 365, by understanding integrations and ensuring our content and preservation systems work together seamlessly.	1	2	3	4	5
DATA QUALITY FOR AI We have specific processes in place to automate the transfer of long-term content to maximize quality for AI use and address missing data, outliers, standardization issues, and potential bias.	1	2	3	4	5
DISCOVERABILITY OPTIMIZATION We enhance content findability through AI-powered enrichment techniques and extend metadata with AI provenance information.	1	2	3	4	5
GOVERNANCE & ACCESS CONTROL We have defined clear permission boundaries between human and AI access and determined what data should not be exposed due to policy, ethics, or proprietary concerns.	1	2	3	4	5
HUMAN OVERSIGHT & ENHANCEMENT We have implemented feedback loops and human validation for sensitive classifications and content enrichment processes.	1	2	3	4	5
PRIVACY & SECURITY PROTECTION We implement PII classification and identification during content ingestion and have processes to manage sensitive content appropriately.	1	2	3	4	5
STRATEGIC VALUE RECOGNITION Our organization views long-term content preservation as a strategic asset rather than just a compliance requirement.	1	2	3	4	5

Assessment Scorecard

Scoring Guide

48-60 points: Optimized

Your organization has a mature digital preservation strategy that not only ensures compliance but positions your long-term content as a strategic asset ready for AI and future innovations.

36-47 points: Managed

Your organization has solid digital preservation practices in place but may have opportunities to further enhance strategic value and AI readiness.

24-35 points: Defined

Basic digital preservation processes exist, but significant improvements could be made in areas of integration, scalability, and strategic value recognition.

12-23 points: Initial

Your organization has recognized the importance of digital preservation but lacks formalized processes and technologies to ensure long-term content usability.

Under 12 points: Ad Hoc

Digital preservation activities are reactive, inconsistent, and your organization faces significant risk of data loss, format obsolescence, and compliance issues.

Assessment Scorecard

Next Steps Recommendation

Based on your assessment score, we recommend:

Optimized (48-60):

Focus on leveraging preserved content for AI/ML initiatives and competitive advantage. Establish centers of excellence for content management and AI integration.

Managed (36-47):

Integrate digital preservation more deeply with your enterprise ecosystem. Enhance metadata enrichment and implement AI-ready content taxonomies.

Defined (24-35):

Implement comprehensive digital preservation technology and establish formal governance policies. Begin addressing AI readiness prerequisites.

Initial (12-23):

Conduct a comprehensive information audit, develop a digital preservation roadmap, and establish baseline content hygiene practices.

Ad Hoc (Under 12):

Secure executive sponsorship, begin risk assessment for critical long-term content, and develop a business case for strategic preservation investment.
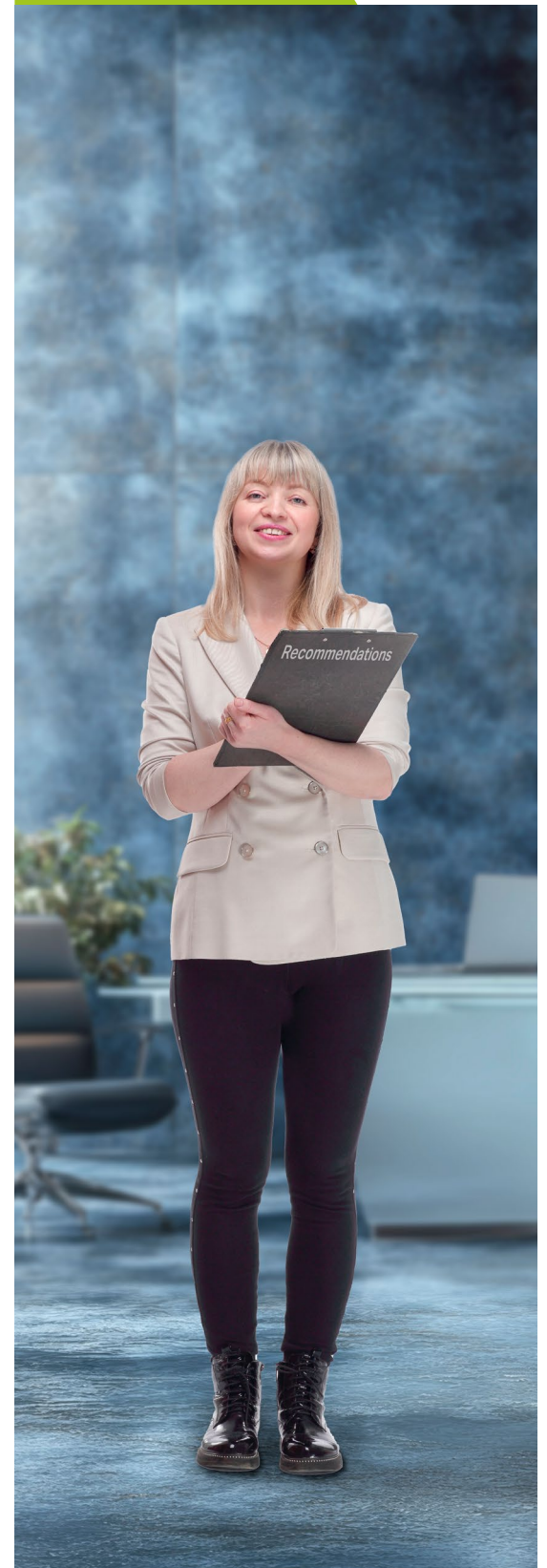
Conclusion

Your long-term content has unprecedented strategic value. Organizations must shift their perspective from viewing long-term content management as merely a compliance burden to recognizing it as a foundation for competitive advantage. By ensuring your long-term content is properly archived, preserved, and optimized for AI use, you simultaneously strengthen your compliance posture and position your organization to create value through AI initiatives.

The convergence of compliance requirements and AI opportunities demands a robust strategy for long-term content - one that addresses both defense (risk mitigation) and offense (value creation). By implementing Active Digital Preservation and leveraging the Microsoft ecosystem, organizations can transform their approach to long-term content and unlock its full potential in the age of AI.

About Preservica

About Preservica

Preservica is changing the way organizations around the world protect and re-use 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.

For more information, visit

Preservica.com

aiim

Association for Intelligent
Information Management

+1 301 587 8202

hello@aiim.org

www.aiim.org

Preservica