



IS YOUR REPOSITORY READY FOR AI?

Why every document repository needs a health check,
and why an active repository needs to be watched around
the clock.

60%

of AI projects unsupported by AI-
ready data will be abandoned
through 2026.

GARTNER, FEBRUARY 2025

59%

of organizations do not measure
data quality at all

52%

of stored data is dark: nobody
knows what it is or what it is
worth

9%

of organizations can scan their
data in real time



01 / THE EVIDENCE

AI IS FAILING ON CONTENT, NOT ON MODELS

Most organizations are not short of information. They hold years, often decades, of valuable content in document repositories. Much of it is hard to find, inconsistently structured, poorly classified, duplicated or out of date. AI does not repair that. It inherits it, and returns confident answers drawn from the wrong version of the truth.

63%

of organizations do not have, or are unsure they have, the data management practices AI requires.²

30%

or more of generative AI projects were forecast to be abandoned after proof of concept. Poor data quality led the list of causes.³

40%

plus of agentic AI projects will be canceled by the end of 2027, on cost, unclear value or inadequate risk controls.⁴

\$12.9M

is the least that poor data quality costs the average organization every year.⁵

WHAT IS ACTUALLY IN THE REPOSITORY

Share of all stored data⁶**52%**

Dark. Unclassified. Its content and value are unknown to the organization.

33%

ROT. Redundant, obsolete or trivial, and known to be useless.

15%

Business critical.

~20%

of the working week is spent looking for internal information, or for a colleague who knows where it is.⁷ An AI assistant pointed at the same repository is searching the same mess, only faster.



02 / WHY 24/7

A HEALTH CHECK IS A SNAPSHOT. YOUR REPOSITORY IS NOT.

Every batch that lands in a live repository changes its condition: new duplicates, missing metadata, superseded versions, permissions that no longer match the content. A score taken in January says little about June. Yet most organizations cannot see their content as it changes, and many cannot see it at all.

241

DAYS

is the global average time to identify and contain a data breach. Problems nobody is watching for stay hidden for most of a year.⁸

\$4.44M

average breach cost, global

\$10.22M

average breach cost, United States

56%

have only partial visibility into where their data is stored⁹

9%

report real-time scanning capability⁹

23%

cannot scan their unstructured data at all⁹

97%

of organizations with an AI-related breach lacked proper AI access controls⁸

	ONE-TIME ASSESSMENT	24/7 MONITORING WITH CTS
What you know	Health on the day of the scan	Health now, re-scored as each batch is ingested
New problems	Found at the next audit, if there is one	An alert the moment a KPI falls below your threshold
Remediation	A list of recommendations	Issues worked in order of risk; routine fixes automatic, judgement calls flagged
Over time	A single baseline that ages	A score you can track and prove



03 / THE HEALTH CHECK

MEASURE BEFORE YOU MODERNIZE

Gartner reports that 59% of organizations do not measure data quality at all.⁵ The AITS Content Tracking System (CTS) closes that gap with a single repository health score: a composite index built from the KPIs you define and weight to your priorities. AI does the measuring, but the measuring stick is yours.

YOU DEFINE
YOUR KPIS

YOU SET
× YOUR WEIGHTS

CTS PRODUCES
**= ONE REPOSITORY
HEALTH SCORE**

EIGHT ASSESSMENT AREAS

01 Content Document types, volumes, age, duplication and obsolete material.	02 Structure Folders, libraries, collections and document types; inconsistencies and gaps.	03 Metadata Availability, consistency and usefulness of classifications and identifiers.	04 Search Barriers that make information hard to locate, retrieve or navigate.
05 Data Quality Duplicate, incomplete, conflicting or potentially obsolete information.	06 Security Access, permissions and sensitivity controls that govern safe use and sharing.	07 Governance Ownership, classification, retention and lifecycle considerations.	08 AI Readiness Whether the environment is a sound foundation for AI search, retrieval and automation.

START WITH THE REPOSITORY YOU ALREADY HAVE

AITS is repository agnostic. The Health Check begins with the environments you already depend on and scores them against your metrics. No platform replacement, no migration.

Enterprise content management
Microsoft SharePoint, OpenText, IBM FileNet, Hyland OnBase, Alfresco

Cloud and specialized platforms
Box, Google Workspace, M-Files, Laserfiche, iManage

File and application stores
Network shares, NAS, Azure Blob, Amazon S3, legacy line-of-business systems



04 / AROUND THE CLOCK

FROM ASSESSMENT TO CONTINUOUS ASSURANCE

The Health Check gives you a scored baseline. CTS keeps it true. As batch processes ingest new content, CTS continues to watch overall health, re-scores the repository and raises an alert whenever a KPI falls below an acceptable threshold. When you are ready, it can remediate as well.

STEP 1

MONITOR

Health is watched continuously as new batches are ingested.

STEP 2

RE-SCORE

The health score is recalculated against your KPIs and weights.

STEP 3

ALERT

You are told whenever a KPI drops below its threshold.

STEP 4

REMEDiate

Issues are worked in order of risk. Routine fixes are automatic; judgement calls go to your team.

A PRACTICAL PATH FROM REPOSITORY TO RESULTS

1 DISCOVER

Understand the repository, its content, structure, processes and operational requirements.

2 ASSESS

Score information quality, metadata, search, governance, security and AI readiness against your KPIs.

3 PRIORITIZE

Identify the improvements with the greatest practical value, without assuming a platform replacement.

4 OPTIMIZE

Implement improvements and introduce automation, analytics and governed AI where they make sense.

FROM REPOSITORY HEALTH TO AI OPPORTUNITY

With a measured, trusted foundation in place, organizations can pursue **AI search** across approved sources, **knowledge assistants**, **document intelligence** that extracts, classifies and summarizes at scale, **metadata enrichment** and **intelligent automation**, all under governed AI controls for security, privacy, access and human oversight.

The caution is simple: do not introduce AI for its own sake. Find where it will deliver measurable business value, and prove it with a score you can track.



FIND OUT YOUR REPOSITORY HEALTH

Your repository may already hold the information needed for the next generation of knowledge management and AI. Book an AITS Repository Health Check to learn how ready it really is, then keep it that way with CTS monitoring.

A SCORED BASELINE

One health score, built on your KPIs, with the detail behind it.

PRIORITIZED ACTIONS

Actionable recommendations, ordered by practical value.

ONGOING ASSURANCE

Monitoring, re-scoring and alerts as your content changes.

CONTACT AITS TO BOOK YOUR REPOSITORY HEALTH CHECK

info@appliedits.com

SOURCES

1. Gartner, "Lack of AI-Ready Data Puts AI Projects at Risk," Feb. 2025.
2. Gartner survey of 1,203 data management leaders, July 2024 (reported Feb. 2025).
3. Gartner, "30% of Generative AI Projects Will Be Abandoned After Proof of Concept by End of 2025," July 2024.
4. Gartner, "Over 40% of Agentic AI Projects Will Be Canceled by End of 2027," June 2025.
5. Gartner, Data Quality research; cost figure from Gartner research, 2020.
6. Veritas, Global Databerg Report (Vanson Bourne; 2,550 respondents, 22 countries), 2016.
7. McKinsey Global Institute, "The Social Economy," 2012.
8. IBM, Cost of a Data Breach Report 2025.
9. Cloud Security Alliance and Thales, "The Rise in Unstructured Data and AI Security Risks," Nov. 2025 (210 IT and security professionals).
Cover figures: 60% from source 1; 59% from 5; 52% from 6; 9% from 9. Monetary figures in USD.