



MARKET RESEARCH

The Market Need for Repository Health

What current research says about AI and unready document content, and where the AITS Content Tracking System (CTS) fits.

60%

of AI projects unsupported by AI-ready data are expected to be abandoned through 2026¹

39%

of organizations say their unstructured content is somewhat or fully ready for AI²

70%

say less than half of their unstructured data is discoverable and usable for AI³

SUMMARY

AI is being funded faster than content is being measured

Organizations are investing in AI faster than they are measuring the content it will read. Structured data has had years of attention. Documents, scans, emails and records have not, and they are what most AI assistants and search tools actually work from.

This page gathers what independent research published in 2025 and 2026 says about that gap, how analysts size the market around it, and what kinds of product already address it. It then sets out where the AITS Content Tracking System (CTS) fits.

The research supports three points

- 1 The need is measured, not assumed.** Independent 2025 and 2026 research agrees that document content is the least AI-ready data organizations hold.
- 2 The market is real and growing.** Analysts size unstructured data management and governance in the billions of dollars, growing 15% to 24% a year.
- 3 One position is still open.** A health score that an executive defines, that covers every repository, and that keeps running after the assessment.

Research current to October 2026. Every figure is linked to its source at the end of the page.

63%

of organizations lack, or are unsure they have, the right data management practices for AI¹

94%

say they struggle to manage unstructured data effectively⁴

40%

delayed an AI assistant rollout by three months or more because of overshared data⁵

01 / THE NEED

What the research says

AI projects fail on data

Gartner predicted in February 2025 that through 2026, organizations will abandon 60% of AI projects that are not supported by AI-ready data. Its survey of data management leaders found that 63% either lack the right data management practices for AI or are unsure whether they have them.¹

Document content is the weak spot

Three separate surveys published in 2026 reach the same conclusion from different directions.

FINDING

RESEARCH

65% say structured data is somewhat or fully ready for AI. Only 39% say the same of unstructured content such as emails, PDFs and images.

Hyland survey of AI decision makers, May 2026²

70% say less than half of their unstructured data is discoverable and usable for analytics or AI.

BARC study of 225 data, analytics and AI leaders, 2026³

94% say they struggle to manage unstructured data effectively.

Nasuni survey of 1,000 enterprise purchasing decision makers, 2026⁴

Buyers already ask for what a health score measures

Asked how they judge whether unstructured data is ready for AI, the leaders in the BARC study named the measures below.³ These are properties of the content itself, and they are the kind of KPI a Repository Health score is built from. The same study found that human validation still ranks as a top readiness measure, which is why CTS keeps people checking the AI's work.

READINESS MEASURE	SHARE NAMING IT
Accuracy	54%
Consistency	48%
Business relevance	44%
Completeness	39%
Timeliness	31%

Rollouts are stalling in practice

A Gartner survey of IT leaders found that overshared data caused 40% of organizations to delay the rollout of a workplace AI assistant by three months or more.⁵ In a separate survey, 86% of organizations had done some content cleanup to prepare for AI, but only 51% had done it across the whole organization.⁶

Cleanup is happening in pieces, and without a baseline to measure it against. An organization that cannot say how healthy its content was before the cleanup cannot show what the cleanup achieved.

02 / THE MARKET

How analysts size it

No analyst sizes content health scoring as a market of its own. It sits inside unstructured data management and governance, which analysts put between roughly 4 billion and 9 billion US dollars in 2025, growing 15% to 24% a year.

MARKET	2025 SIZE (USD)	FORECAST (USD)	ANNUAL GROWTH
Unstructured data management ⁷	9.2 billion	28.5 billion by 2033	15.3%
Unstructured data governance ⁸	3.79 billion	10.99 billion by 2030	23.8%
Unstructured data solutions, the broadest definition, including storage and analytics ⁹	Not stated in the public summary	Adds 39.68 billion between 2025 and 2030	15.2%

Figures are from the public summaries of each firm's report. The firms draw the market boundary differently, so the rows are not additive.

These figures show direction more than precision. What they agree on is the rate of growth, and what is driving it. The governance segment is the closest match to repository health and the fastest growing, and the report that sizes it names unstructured data quality and AI readiness as its leading trend.⁸

03 / WHAT IS ON OFFER

Four kinds of answer, each with a limit

Scoring content for AI readiness is no longer a new idea. The offerings available today fall into four types, and each is tied to one platform, one audience or one moment.

TYPE OF OFFERING	WHAT IT DOES	WHAT IT LEAVES OPEN
Platform-native readiness checks	Built into a single collaboration platform. They report oversharing and readiness for that platform's own AI assistant, and some give one readiness score.	One platform only. The vendor sets the weights, not the client.
Discovery and cleanup tools	Find redundant, obsolete and sensitive content across file stores, usually ahead of a migration or a cleanup.	Built around a cleanup project, not an executive's ongoing measure of health.
Data quality platforms adding documents	Monitor continuously and score individual documents before they feed an AI pipeline.	Built for data teams and data lakes, not for records repositories.
Consulting-led assessments	A one-time AI readiness review, delivered as a report.	A point in time. Nothing re-scores the content as it changes.

A single score is therefore not what sets one product apart from another. What matters is whose score it is, how much of the organization's content it covers, and whether it is still true next month.

04 / WHERE CTS FITS

The executive's own score, for every repository, kept current

CTS gives an executive a single Repository Health score: an at-a-glance answer to a question that used to take a sit-down with BI charts and reports. A high score is grounds for confidence. A low one means there is work to do, and the breakdown behind the score says what that work is.

Four things set it apart

- 1 The client defines and weights the KPIs.** Other scores use the vendor's formula. A composite built from the measures an executive already uses to judge health is easier to trust and to defend.
- 2 It works with any repository.** Most organizations hold content in several systems, and more than half of the databases holding unstructured data are on-premises or hybrid. A check tied to one platform cannot see the rest.³

- 3 It is continuous, and it learns.** When staff confirm, correct or dismiss a finding, that decision sharpens the model. A one-time assessment stops at the report.
- 4 It suits records of legal weight.** Scanned, historical and registry records are harder than office documents. Even a national land registry reports that handwritten and poorly scanned records remain difficult for automated systems.¹⁴

The official repository is never touched

CTS runs on a replica of the organization's environment in a secure cloud environment hosted in-country. Nothing it does is destructive, and any result can be compared against the original. That answers the residency and risk questions public-sector buyers raise first.

People stay in charge

CTS recommends action on each issue it finds. Once it has learned how the organization resolves a type of issue, it can be given permission to handle future issues of the same kind, with staff checking its work.

WHOSE SCORE

Built from KPIs the executive and managers choose. AITS recommends a starting set.

HOW MUCH IT COVERS

Any repository, on-premises or cloud, from zero to all of the content.

HOW LONG IT HOLDS

Re-scored as content arrives, with an alert when a KPI falls below its threshold.

Where the need is sharpest

The pressure is greatest where records carry legal weight and reach back decades. Modernization money is being spent in these areas now, mostly on systems and channels.

Land titles	Registries are moving from paper to online systems of record, with electronic submission and new legislation to support it. Those programs fund systems and channels. The measured health of the records behind them is usually not in scope, and a new system that fills forms from existing title data makes the quality of that data visible.
Vital events	Demand on historical records is rising. One Canadian province put more than one million historical vital statistics records online in 2026 and reported a surge in requests from people needing documents to support citizenship applications. Many of those records began on paper and film. ¹¹
Human resources	The driver is exposure. An AI assistant surfaces whatever a user can technically reach, and HR records and pay details are among the first things to appear when permissions were never cleaned up. ¹³

Proof that AI works on registry records, with people checking it

HM Land Registry reported in March 2026 that its AI tool processed one local authority's land charges records in four weeks with four staff, against an estimated three months with 20. Every batch passed its data quality checks the first time, and a quality assurance team still verifies the output.^{10 14}

Policy is pointing the same way

Canada's AI Strategy for the Federal Public Service 2025-2027 sets up an AI Centre of Expertise whose stated role includes supporting data readiness.¹² Readiness of the underlying information is now part of how government expects AI to be adopted.

AITs WHITE PAPERS AT APPLIEDITS.COM

Land Titles Registry | Vital Events Registry | Human Resources

Find out your repository health

A Repository Health Check gives an executive a scored baseline, the detail behind it, and a prioritized list of what to do next. CTS monitoring keeps that baseline true as content changes.

Contact AITS to book a Repository Health Check: info@appliedits.com

ABOUT AITS

Applied IT Solutions (AITS) is a Toronto-based consultancy founded in 2008 and led by its principal consultant. It advises government and Fortune 500 organizations on artificial intelligence, enterprise data strategy, document and records management, and business process improvement, with one aim: deriving tangible business value from the application of information technology. Most engagements are delivered end to end by the principal, from requirements gathering through solution design and delivery to train-the-trainer and hand-off to the client's own team. For larger projects, AITS brings in experienced consultants it works with regularly. AITS has run projects worldwide and works in any time zone. Out of this work, AITS developed the Content Tracking System (CTS), an AI-powered monitor that scores the health of a document repository against the organization's own KPIs.

SOURCES

1. Gartner, "Lack of AI-Ready Data Puts AI Projects at Risk," Feb. 2025; survey of data management leaders, July 2024.
<https://www.gartner.com/en/newsroom/press-releases/2025-02-26-lack-of-ai-ready-data-puts-ai-projects-at-risk>
2. IT Brief, "Enterprise AI ambitions outpace readiness, survey finds," May 2026; reporting Hyland research.
<https://itbrief.co.uk/story/enterprise-ai-ambitions-outpace-readiness-survey-finds>
3. IDM, "Most Unstructured Data Not Ready for AI, Study Finds," 2026; reporting the BARC study "Harnessing Unstructured Data for AI Innovation."
<https://idm.net.au/node/15659>
4. Unite.AI, coverage of Nasuni, "The State of Enterprise File Data Annual Report 2026."
<https://www.unite.ai/?p=423340>
5. IDM, "Copilot Gets Sensitive Data Guardrails in Purview," 2026; reporting a Gartner survey of IT leaders.
<https://idm.net.au/node/15542>
6. ShareGate, survey of IT leaders on AI governance and content cleanup, 2026.
<https://sharegate.com/blog/93-of-it-leaders-are-confident-in-their-ai-governance-but-nearly-1-in-3-report-data-exposure-incidents>
7. DataHorizon Research, "Unstructured Data Management Market," forecast 2026 to 2033.
<https://datahorizonresearch.com/unstructured-data-management-market-71856>
8. The Business Research Company, "Unstructured Data Governance Market Report 2026."
<https://www.thebusinessresearchcompany.com/report/unstructured-data-governance-market-report>
9. Technavio, "Unstructured Data Solution Market Analysis, Size, and Forecast 2026-2030."
<https://www.technavio.com/report/unstructured-data-solution-market-industry-analysis>
10. HM Land Registry, "Transforming local land charges migrations with AI," March 2026.
<https://hmlandregistry.blog.gov.uk/2026/03/17/transforming-local-land-charges-migrations-with-ai>
11. Government of New Brunswick, "Provincial Archives putting history within easy reach," 2026.
<https://www.gnb.ca/en/news/n-b.2026.07.provincial-archives-putting-history-within-easy-reach.html>
12. Canadian Government Executive, summary of the AI Strategy for the Federal Public Service 2025-2027.
<https://canadiangovernmentexecutive.ca/?p=15682>
13. CoreView, "Best Practice for a Microsoft Copilot Readiness Assessment," Aug. 2026.
<https://www.coreview.com/blog/best-practice-for-a-microsoft-copilot-readiness-assessment>
14. Results Sense, "HM Land Registry AI tool cuts land charges migration from months to weeks," March 2026.
<https://www.resultsense.com/news/2026-03-23-hm-land-registry-ai-speeds-local-land-charges-migration/>

Market figures are in US dollars. Survey figures are as reported by each source.