



THE EMPLOYEE FILE IS A SYSTEM OF RECORD

Measuring and sustaining document repository health in human resources, before AI and after it.

70%

of HR time goes to administrative tasks in one 2026 survey of smaller employers

46%

of data breaches studied in 2025 involved employee personal data

88%

of HR leaders say AI tools have not yet delivered significant business value

EXECUTIVE SUMMARY

EVERY EMPLOYMENT DECISION LEAVES A DOCUMENT

HR runs on documents: offers and contracts, eligibility and identity checks, policy acknowledgements, performance reviews, leave and accommodation records, disciplinary files, payroll and benefits forms, and separation papers. Together they are the evidence of how every employee was hired, paid, managed and let go. When they are missing, duplicated or kept too long, the organization cannot prove what it did.

Most organizations hold that evidence in several places at once: HR applications, document repositories, spreadsheets, workflow systems and historical records. HR is now under pressure to modernize, with self-service, automation and AI. All of it depends on the condition of the repository underneath.

Yet few HR departments can state, with evidence, how complete, consistent and compliant their employee records are today, or whether that condition is improving or declining.

THIS PAPER ARGUES THREE THINGS

- 1 Repository condition is the limiting factor.** It decides how far HR can modernize, and it decides whether AI helps or harms.
- 2 Condition can be measured.** A single health score, built from KPIs the organization defines and weights, replaces opinion with evidence.
- 3 Measurement has to be continuous.** A workforce changes every working day, so a one-time assessment starts to age as soon as it is finished.

The paper then sets out how the AITS Content Tracking System (CTS) does this, with a worked example of the score, an illustrative scenario, and what a Repository Health Check involves.

70%

of HR time goes to administrative tasks in a 2026 survey of Canadian small and mid-sized employers ¹

46%

of data breaches studied in 2025 involved employee personal data ⁴

63%

of organizations lack, or are unsure they have, the data practices AI requires ⁷



01 / THE INHERITANCE

ONE EMPLOYEE, A DOZEN PLACES TO LOOK

An employee file is unlike most business records. It begins before the person is hired and has to outlast their departure, sometimes by years. Recruiters, managers, payroll, benefits providers and the employee all add to it, usually in different systems. And parts of it, such as medical and accommodation records, must be held apart from the rest.

The layers build up: paper files scanned in batches; documents attached to an HR information system; folders on shared drives named by whoever created them; forms that arrived by email; records migrated from a previous system or an acquired company. Each has its own conventions and its own gaps.

The work this creates is measurable. In a 2026 survey of more than 450 HR professionals at Canadian small and mid-sized employers, 70% of HR time went to administrative tasks, with managing employee files and HR administration the largest share at 37%.¹

WHAT THIS PRODUCES IN PRACTICE

- Employee files with required documents missing, unsigned or expired.
- The same document held in three places, with no record of which is authoritative.
- Records for one person split across systems under different identifiers.
- Medical or disciplinary records stored where too many people can see them.
- Files of former employees kept long past their retention period, or deleted before it.

None of this is unusual. It is what years of hiring, restructuring and system changes look like. The question is whether anyone can say how much of each there is.

18%

of the employers surveyed have fully integrated HR systems¹

THE EMPLOYEE LIFECYCLE

Recruit, hire, onboard, pay, develop, manage, separate, retain, dispose. Every stage adds documents, and every stage has its own rules.



02 / THE STAKES

WHY HR CANNOT TREAT THIS AS HOUSEKEEPING

Employee records are regulated twice. Employment law says what must be kept and for how long. Privacy law says what must be protected and when it must go. A repository that cannot show both is exposed in both directions.

KEEPING TOO LITTLE IS PENALIZED

In the United States, for example, employers must keep personnel records for at least a year and payroll records for three, and must preserve everything relevant once a discrimination charge is filed.² A single incomplete employment eligibility form carries a penalty of \$288 to \$2,861, and 2026 guidance made more than ten kinds of error immediately finable. For 200 flawed forms that is roughly \$57,600 to \$572,200.³

KEEPING TOO MUCH IS A LIABILITY

Employee personal data was involved in 46% of the data breaches studied in 2025, up from 37% a year earlier, at an average cost of \$168 per record.⁴ Every file held past its retention date is exposure with nothing to offset it.

DISPUTES TURN ON DOCUMENTS

A grievance, a dismissal claim or an access request is answered from the file. If the signed policy, the warning letter or the accommodation record cannot be produced, the organization's account of events is unsupported.

THE STANDARD IS ALREADY WRITTEN

ISO 15489 says an authoritative record has four characteristics: authenticity, reliability, integrity and usability.⁵ HR meets its obligations when it can show these hold across its files, not only assert that they do.

\$2,861

maximum US penalty for one incomplete employment eligibility form³

\$168

average cost per employee record exposed in a breach⁴

AN AUTHORITY
RECORD IS

Authentic: it is what it claims to be.

Reliable: its contents can be trusted.

Complete and unaltered: it has integrity.

Usable: it can be found, read and understood.



03 / AI

AI RAISES THE STAKES

HR has good reasons to look at AI. The useful applications are concrete: document classification, information extraction, OCR and interpretation of historical records, record discovery and search, duplicate and anomaly identification, metadata enrichment, workforce reporting and staff knowledge assistance.

The results so far are sobering. In a 2025 Gartner survey, 88% of HR leaders said their organizations had not realized significant business value from AI tools.⁶

WHAT STOPS AI IS CONSISTENT

Gartner also predicts that through 2026 organizations will abandon 60% of AI projects that are not supported by AI-ready data, and reports that 63% either lack the right data management practices for AI or are unsure whether they have them.⁷

IN HR THE FAILURE IS SPECIFIC

- AI does not know which of three copies of a contract is the signed one.
- It will answer a policy question from a superseded version if nothing marks it as superseded.
- It will surface a medical or disciplinary record to someone who should never see it, if the permissions are wrong.
- It will summarize a file that is missing the document that mattered.

Elsewhere that produces a wrong answer. In HR it can produce a wrong decision about a person.

“AI becomes more useful when HR information has already been digitized, structured and governed.”

88%

of HR leaders say AI tools have not yet delivered significant business value⁶

60%

of AI projects without AI-ready data expected to be abandoned through 2026⁷

63%

of organizations lack, or are unsure they have, the data practices AI requires⁷



04 / DEFINITION

WHAT “HEALTHY” MEANS FOR AN HR REPOSITORY

Health is not one property. CTS evaluates a repository across eight dimensions. The table below translates each into the question an HR leader would ask.

DIMENSION	THE HR QUESTION
Completeness	Does every employee file hold the documents its role, location and employment type require?
Metadata quality	Are employee identifiers, document types, dates and signatures present, valid and consistent across systems?
Compliance	Do holdings meet employment standards, eligibility, health and safety, and privacy requirements in each jurisdiction?
Retention	Is each record class held for its required period and no longer, with legal holds honoured?
Security	Do access permissions separate medical, disciplinary and investigation records from the general file?
Governance	Does every collection have an owner, a classification and a recorded source?
Currency	Can the current contract, policy acknowledgement or certification be told apart from superseded or expired ones?
Duplication	Where the same document exists more than once, is the authoritative copy identified?

These are starting points. The KPIs beneath each dimension are the organization’s own, because only it knows which document types carry the most risk, which jurisdictions it answers to, and which gaps its auditors will not tolerate.

Most organizations never take this step. Gartner reports that 59% do not measure data quality at all.⁸ Without a measure there is no baseline, no trend, and no way to show an auditor, a regulator or a tribunal that the record is sound.



05 / METHOD

HOW THE SCORE WORKS

CTS produces a Repository Health score: a composite of component indices, each measuring one KPI. Three decisions belong to the organization.

DECISION 1**WHAT TO MEASURE**

The KPIs, expressed in HR's own terms.

DECISION 2**HOW MUCH IT MATTERS**

A weight for each KPI, reflecting legal and operational risk.

DECISION 3**WHAT IS ACCEPTABLE**

A threshold for each KPI, below which an alert is raised.

The AI is trained on the organization's own repository to recognize each KPI. The measuring is automated. The standard is not.

A WORKED EXAMPLE

KPI	WEIGHT	SCORE	CONTRIBUTION
File completeness	30%	76	22.8
Retention conformance	25%	70	17.5
Access control conformance	20%	88	17.6
Document currency	15%	82	12.3
Duplicate resolution	10%	64	6.4
Repository Health score	100%		76.6

Illustrative figures to show the arithmetic. They are not client data. Contribution = weight × score.

READING THE SCORE

The composite says how confident to be. The components say where to act. Here duplicate resolution is the lowest component, yet retention conformance and file completeness cost the composite more, because they carry more weight. Weighting directs effort to what matters most, not to what looks worst.

The same score can be produced for the whole repository or for any section of it: a business unit, a jurisdiction, or a single document type.



06 / CONTINUITY

FROM BASELINE TO CONTINUOUS ASSURANCE

A Health Check is a snapshot. A workforce is not. People are hired, moved and separated every working day, and policies, certifications and retention clocks change the repository further. A baseline measured once starts to age as soon as it is finished.

Few organizations can watch their own content change. In a 2025 Cloud Security Alliance survey, 56% had only partial visibility into where their data is stored, only 9% reported real-time scanning, and 23% could not scan their unstructured data at all.⁹

STEP 1**MONITOR**

Content is watched as it is added, changed, migrated, classified, retained or removed.

STEP 2**RE-SCORE**

The health score is recalculated against the organization's KPIs and weights.

STEP 3**ALERT**

An alert is raised whenever a KPI falls below its threshold.

STEP 4**REMEDiate**

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

WHAT AN ALERT LOOKS LIKE

ILLUSTRATIVE ALERT, NOT CLIENT DATA

KPI	File completeness
Scope	Employees onboarded in the last 30 days
Threshold / measured	95 / 88
Likely cause	Signed policy acknowledgement missing at one location
Recommended action	Route affected files to HR operations; check the onboarding checklist

ASSESSMENT AND MONITORING COMPARED

	ONE-TIME ASSESSMENT	24/7 MONITORING WITH CTS
What you know	Health on the day of the scan	Health now, re-scored with each change
New problems	Found at the next audit, if any	An alert when a KPI falls below threshold
Over time	A single baseline that ages	A trend you can report and defend



07 / SCENARIO

AN EMPLOYER THAT HAS GROWN BY ADDITION

This is a composite scenario, not a client account. It describes how a Health Check and monitoring would proceed in an organization of the kind this paper describes.

An organization has grown through hiring, restructuring and two acquisitions. Employee records sit in an HR system, a document repository, shared drives and scanned paper. Leadership wants an AI assistant for HR and asks a fair question: are our records ready? Nobody can answer with evidence.

- | | |
|---------------------|---|
| 1 DEFINE | HR, legal, privacy and records staff agree the KPIs, weights and thresholds. File completeness and retention conformance weigh most. |
| 2 BASELINE | CTS scores the whole repository, each business unit and each document type. The likely pattern: recent hires score highest, acquired and long-serving employees lowest. |
| 3 PRIORITIZE | Findings are ranked by risk, not volume. Typical items: former employees' files past retention, and sensitive records in open folders. |
| 4 DECIDE | The AI pilot is scoped to the sections whose scores justify it. Remediation targets the sections that would mislead it. |
| 5 MONITOR | CTS stays on. The score is reported to leadership as a trend, and each hire, change and separation is checked as it happens. |

WHO IS IN THE ROOM

HR operations, legal or employee relations, privacy, records management, and IT staff who know the systems of record.

WHAT IS NOT REQUIRED

A new HR system, a migration, or a decision about AI. The Health Check measures what is already there.

WHAT TO EXPECT

Uneven scores. Files built over years and across acquisitions will not be uniform, and the variation is the useful part.

WHAT CHANGES

The conversation moves from “we believe the files are in reasonable shape” to a number, a trend and a ranked list. HR can show an auditor how it knows.



WHAT A REPOSITORY HEALTH CHECK INVOLVES

The engagement begins with an AI Health Check: a structured assessment of an existing repository that establishes the first Repository Health baseline and identifies information and governance risks.

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 the organization's 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.

WHAT HR PROVIDES

- Read access to the repository, or to representative sections of it.
- The employment, privacy and retention rules the holdings must meet.
- People who know the files, to define the KPIs and weights.

WHAT HR RECEIVES

- A baseline health score, overall and by section.
- The component breakdown behind it.
- Identified information and governance risks.
- Prioritized, actionable recommendations.

WHERE IT SITS IN MODERNIZATION

AITS describes HR modernization as an eight-step pathway. The health score is not a ninth step. It runs alongside all eight and shows whether each one left the repository better than it found it.



REPOSITORY HEALTH SCORE, MEASURED CONTINUOUSLY ACROSS EVERY STEP

CTS works with any repository. Modernization can begin with existing records and processes: the Health Check starts with the systems HR already depends on and does not presuppose a migration.



FIND OUT THE HEALTH OF YOUR EMPLOYEE RECORDS

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

CONTACT AITS TO BOOK A REPOSITORY HEALTH CHECK

info@appliedits.com

ABOUT AITS

Applied IT Solutions (AITS) is the deployment group of Applied IT, based in Toronto, Canada, with a global support program. Its team brings more than 100 years of combined consulting experience in enterprise information management, document and records repositories, business process improvement and technology deployment, serving government, banking, insurance and manufacturing. AITS configures commercial software to fit how information moves through an organization, so that the technology supports the process and the process does not have to work around the technology.

SOURCES

1. Folks, survey of more than 450 HR professionals at Canadian small and mid-sized businesses, April 2026.
2. US Equal Employment Opportunity Commission, Recordkeeping Requirements.
3. Morgan Lewis, "ICE Rewrites the Rules on Form I-9 Violations," April 2026.
4. IBM, Cost of a Data Breach Report 2025.
5. ISO 15489-1:2016, Information and documentation: Records management, Part 1.
6. Gartner, survey of 114 HR leaders, July 2025 (released Oct. 2025).
7. Gartner, "Lack of AI-Ready Data Puts AI Projects at Risk," Feb. 2025; survey of 1,203 data management leaders, July 2024.
8. Gartner, Data Quality research.
9. Cloud Security Alliance and Thales, "The Rise in Unstructured Data and AI Security Risks," Nov. 2025. Monetary figures are in US dollars.