



EVERY IDENTITY BEGINS WITH A RECORD

Measuring and sustaining document repository health in vital events registries, before AI and after it.

150M

children under five worldwide whose births have never been registered

60%

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

18.9M

people born in 1920 or earlier with no death recorded in one national file



EXECUTIVE SUMMARY

A REGISTRY OF LIVES, AND EVERY SERVICE DEPENDS ON IT

A vital events registry records the facts of a person's legal existence: birth, marriage, change of name, adoption, death. Each registration becomes the source from which other things are issued: a certificate, a passport, a pension, a settled estate. When the record is right, nobody notices. When it is wrong, missing or duplicated, a person cannot prove who they are.

Most registries hold these records across generations of technology: bound registers, paper registrations, microfilm, scanned images and electronic registrations. They are now under pressure to modernize, with online ordering, data sharing with other programs, and AI for indexing, matching and search. All of it depends on the condition of the repository underneath.

Yet few registries can state, with evidence, how complete, consistent and trustworthy their holdings 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 a registry can modernize, and it decides whether AI helps or harms.
- 2 Condition can be measured.** A single health score, built from KPIs the registry defines and weights, replaces opinion with evidence.
- 3 Measurement has to be continuous.** An active registry 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.

150M

children under five worldwide whose births are unregistered³

14,000

versions of the birth certificate found in circulation in one country⁵

63%

of organizations lack, or are unsure they have, the data practices AI requires¹⁰

01 / THE INHERITANCE

A LIFETIME OF RECORDS, KEPT LONGER THAN A LIFETIME

A vital events repository is unlike most enterprise content stores. Its records have to stay accessible, accurate and protected for generations. A birth registered ninety years ago may be requested tomorrow to settle a pension, and a death record closes files across government. Nothing ages out.

The United Nations lists the events a civil registration system records: live births, deaths and foetal deaths; marriages, divorces, annulments and separations; adoptions, legitimations and recognitions.¹ Each has its own form, its own informants and its own history of amendment.

That history arrives in layers: bound registers and paper registrations; microfilm and fiche; scanned images from several campaigns; index data keyed by hand and migrated between systems; and electronic registrations. One state review in 2024 found that most records in its vital records office were still paper, many of the oldest in fragile condition, with digitization on hold pending funding.²

WHAT THIS PRODUCES IN PRACTICE

- Registrations that exist as images but carry incomplete or inconsistent index data.
- One person recorded under variant spellings, with nothing linking the records.
- Amendments, corrections and adoptions held apart from the registration they change.
- A birth with no link to the later marriage, change of name or death of the same person.
- Handwritten registers a person can read but OCR cannot.

None of this is unusual. The question is whether anyone can say how much of each there is.

45%

growth in record transactions at one state vital records office, 2015–17 to 2021–23²

A RECORD FOR EVERY STAGE

Birth, adoption, change of name, marriage, divorce and death are separate registrations. They describe one person only if they are linked.



02 / THE STAKES

WHY A REGISTRY CANNOT TREAT THIS AS HOUSEKEEPING

A birth certificate is a breeder document: it lets the holder obtain other documents and benefits, such as a passport, a driver's licence or a pension.⁵ An error or a gap in the registry therefore travels. It is copied into every system that relied on the certificate.

THE INTEGRITY RISK IS DOCUMENTED

A United States federal review found more than 14,000 different versions of the birth certificate in circulation, issued by 6,422 different entities, and identified that variety as an opening for fraud.⁵ Verifying a certificate depends on being able to go back to a trustworthy registration.

A MISSING EVENT HAS A COST

Worldwide, nearly 40% of deaths are never registered.⁴ When a death is not recorded against a person, their identity stays open. A 2023 audit of the US Social Security number file found about 18.9 million people born in 1920 or earlier with no death information on record, a gap equal to more than 10% of the federal death file.⁶

TIMELINESS IS MEASURABLE, AND OFTEN MISSED

A 2024 state audit found that 84% of fetal death registrations took longer than the eight days the law allowed, averaging 26 days, and that the system could not notify anyone when a certificate needed action.⁷

THE STANDARD IS ALREADY WRITTEN

ISO 15489 says an authoritative record has four characteristics: authenticity, reliability, integrity and usability.⁸ A registry meets its mandate when it can show these hold across its holdings, not only assert that they do.

18.9M

people born in 1920 or earlier with no death information in one national file⁶

84%

of one state's fetal death registrations missed the legal deadline⁷

AN AUTHORITATIVE
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

Registries have 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, and comparison of amendments against the record.

Gartner predicts that by 2029, 60% of government agencies will use AI agents to automate more than half of citizen transactional interactions, up from fewer than 10% in 2025.⁹

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.¹⁰ In government specifically, a survey of 492 leaders in 14 countries found that only 26% had integrated AI across their organization, and 45% reported inadequate data infrastructure.¹¹

IN A REGISTRY THE FAILURE IS SPECIFIC

- AI does not know which of two registrations is authoritative.
- It will read a name from a faded register and present it with confidence.
- It will merge two people who share a name, or miss one person recorded twice.
- It will answer from a registration that a later amendment has superseded.

Elsewhere that produces a wrong answer. In a registry it can produce a wrong answer about who a person is.

Information structure comes first. Establish an ongoing healthy document repository, then choose the systems that sit on top of it.

60%

of AI projects without AI-ready data expected to be abandoned through 2026¹⁰

26%

of government organizations surveyed have integrated AI across the organization¹¹

45%

of them report inadequate data infrastructure¹¹



04 / DEFINITION

WHAT “HEALTHY” MEANS FOR A VITAL EVENTS REPOSITORY

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

DIMENSION	THE REGISTRY QUESTION
Completeness	Does every registered event have its registration image, its index record and its supporting documents?
Metadata quality	Are names, dates, places and registration numbers present, valid and consistent across eras?
Compliance	Do holdings meet the vital statistics statute, the registrar’s directions, and privacy and records legislation?
Retention	Is each record class held for its required period, and is transfer to the archives recorded when it happens?
Security	Do access permissions reflect the sensitivity of the record, including adoptions and sealed registrations?
Governance	Does every collection have an owner, a classification and a recorded source?
Currency	Can the current registration be told apart from one that an amendment, correction or adoption has superseded?
Duplication	Where the same event is recorded more than once, is the authoritative registration identified?

These are starting points. The KPIs beneath each dimension are the registry’s own, because only the registry knows which event types carry the most risk, which eras were captured under which rules, and which gaps its statute will not tolerate.

Most organizations never take this step. Gartner reports that 59% do not measure data quality at all.¹² Without a measure there is no baseline, no trend, and no way to show a minister, an auditor or the public 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 registry.

DECISION 1**WHAT TO MEASURE**

The KPIs, expressed in the registry's own terms.

DECISION 2**HOW MUCH IT MATTERS**

A weight for each KPI, reflecting statutory 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 registry's own repository to recognize each KPI in the content. The measuring is automated. The standard being measured against is not.

A WORKED EXAMPLE

KPI	WEIGHT	SCORE	CONTRIBUTION
Record-linkage integrity	30%	78	23.4
Index completeness	25%	74	18.5
Image legibility for OCR	15%	66	9.9
Statutory compliance	15%	92	13.8
Access control conformance	15%	90	13.5
Repository Health score	100%		79.1

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. In this example image legibility is the lowest component, yet record-linkage integrity and index completeness cost the composite more, because they carry more weight. Weighting directs effort to what matters most to the registry, not to what looks worst.

The same score can be produced for the whole repository or for any section of it: an era, an event type, or a single capture campaign.



06 / CONTINUITY

FROM BASELINE TO CONTINUOUS ASSURANCE

A Health Check is a snapshot. An active registry is not. Events are registered every working day, and amendments, corrections, conversions and transfers 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.¹³

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 registry'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	Index completeness
Scope	Death registrations ingested in the latest batch
Threshold / measured	90 / 84
Likely cause	Date of birth missing from one submission source
Recommended action	Route affected registrations to indexing review; check intake validation

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 batch
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

A REGISTRY THAT HAS TO REMEMBER EVERYONE

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

A registry holds records from bound registers through to electronic registration. Leadership wants AI-assisted search and matching, and asks a fair question: are our records ready? Nobody can answer with evidence.

- | | |
|---------------------|---|
| 1 DEFINE | The registrar, records staff and registration officers agree the KPIs, weights and thresholds. Record linkage and index completeness weigh most. |
| 2 BASELINE | CTS scores the whole repository, each era and each event type. The likely pattern: recent electronic registrations score highest, handwritten registers lowest. |
| 3 PRIORITIZE | Findings are ranked by risk, not volume. Typical items: amendments not linked to the registration they change, and deaths with no birth record linked. |
| 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 new batch is checked as it arrives. |

WHAT CHANGES

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

WHO IS IN THE ROOM

The registrar or delegate, records management, registration officers who know the event types, privacy staff, and IT staff who know the systems of record.

WHAT IS NOT REQUIRED

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

WHAT TO EXPECT

Uneven scores. A repository built over generations will not be uniform, and the variation is the useful part.



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 registry'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 THE REGISTRY PROVIDES

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

WHAT THE REGISTRY 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 registry 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. The Health Check starts with the systems the registry already depends on and does not presuppose a migration.



FIND OUT WHAT YOUR REGISTRY RESTS ON

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

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 the technology supports the process.

SOURCES

1. United Nations, Principles and Recommendations for a Vital Statistics System, Revision 3, 2014.
2. Rhode Island Office of Management and Budget, Performance Management Review: Office of Vital Records, Nov. 2024.
3. UNICEF, The Right Start in Life: Global Levels and Trends in Birth Registration, 2024 update, Dec. 2024.
4. World Health Organization, "Civil Registration and Vital Statistics" fact sheet, Feb. 2024.
5. US Department of Health and Human Services, Office of Inspector General, Birth Certificate Fraud, OEI-07-99-00570, Sept. 2000.
6. US Social Security Administration, Office of the Inspector General, audit A-06-21-51022, July 2023.
7. California State Auditor, report 2023-110, Feb. 2024.
8. ISO 15489-1:2016, Information and documentation: Records management, Part 1.
9. Gartner, "Gartner Reveals Top Technologies Shaping Government AI Adoption," Sept. 2025.
10. Gartner, "Lack of AI-Ready Data Puts AI Projects at Risk," Feb. 2025; survey of 1,203 data management leaders, July 2024.
11. EY, survey of 492 government leaders in 14 countries, conducted Aug.–Sept. 2024, released June 2025.
12. Gartner, Data Quality research.
13. Cloud Security Alliance and Thales, "The Rise in Unstructured Data and AI Security Risks," Nov. 2025.