AJA Labs.ai

Data is the New Working Paper

Playing with Data, Analytics and Decision Intelligence

CA Anand Prakash Jangid
CA, CISA | Founder & Chief Change Agent, AJALabs.ai

A STORY TO BEGIN WITH

Washington DC, 26 June 2023

While our profession was still debating whether data analytics was worth the effort, a regulator eleven thousand kilometres away published an exposure draft that quietly assumed we were already doing it.

PCAOB Release No. 2023-004

Proposed Amendments Related to Aspects of Designing and Performing Audit Procedures that Involve Technology-Assisted Analysis of Information in Electronic Form.

Docket 052. Comments invited until 28 August 2023. It proposed to amend AS 1105, Audit Evidence, and AS 2301, The Auditor's Responses to the Risks of Material Misstatement — provisions that had not been substantially changed since 2010.

Read the title again

- It does not ask whether auditors should use data analysis.
- It asks how the standards should govern the analysis auditors are already performing.
- The debate we are having in 2026 was already over in 2023 — somewhere else.

Source: PCAOB, Release No. 2023-004 (26 June 2023), Rulemaking Docket Matter No. 052.

**If your working paper cannot be re-run,
it is not evidence.
It is a souvenir.**

Hold that thought for the next fifty minutes.

THE NEXT FIFTY MINUTES

Five movements

01	**What is changing**	The client, the evidence and the regulator all moved. Our method did not.
02	**Oil without a refinery**	The most misquoted line in our profession — and what it actually means.
03	**Dark data**	The 55–80% of enterprise data nobody audits, and where intent actually hides.
04	**Why data-driven wins**	The evidence: PCAOB, ACFE 2026, NFRA.
05	**Monday morning**	A maturity ladder, six tests

MOVEMENT ONE

What is changing around us

Three shifts, all at once, and none of them optional

MOVEMENT 1

Three shifts have already happened

1

The transaction changed

It is no longer typed. It is triggered — by an API, a bot, a scanner, a rule. Nobody signs it. A log records it.

2

The evidence changed

The signed voucher is now a workflow timestamp, an approver ID, an IP address and a change-log entry.

3

The expectation changed

Boards want assurance continuously, not ninety days after year-end. Regulators want the population, not your sample.

Our audit method was designed for a world where obtaining evidence was expensive. That world is gone.

THE INCONVENIENT ORIGIN STORY

Sampling was a technology of scarcity

Why we sample

- Ledgers were paper. Vouchers lived in godowns.
- Testing 100% was physically impossible.
- So the profession made a brilliant trade: accept sampling risk, gain feasibility.
- SA 530 codified that trade-off. It never mandated it.

What changed

- The full population now arrives as a file, in minutes, for free.
- The constraint moved from access to capability.
- Sampling risk is now a choice we make, not a limit we inherit.
- And a choice must be justified to a regulator.

Question for the room: if you can test one hundred per cent, what exactly are you sampling for?

DO THE ARITHMETIC WITH ME

The coverage we actually achieve

12,00,000

journal lines in a mid-size manufacturer's financial year

60

entries typically selected for journal entry testing

0.005%

of the population actually examined

99.995%

never seen by any auditor, ever

We are not bad auditors. We are auditors running a nineteenth-century sampling economy on twenty-first-century data volumes.

Illustrative volumes based on typical mid-market ERP populations — calibrate to your own client base.

BACK TO THAT EXPOSURE DRAFT

Three years, and it is already law somewhere

26 Jun 2023

Exposure draft

PCAOB Release 2023-004 issued for comment. Docket 052. Comments close 28 August 2023.

12 Jun 2024

Adopted

PCAOB Release 2024-007 amends AS 1105, Audit Evidence, and AS 2301, Auditor's Responses to Risks.

20 Aug 2024

SEC approves

SEC Release 34-100774. Approved unanimously by the Board, alongside AS 1000.

15 Dec 2025

In force

Effective for audits of financial statements for fiscal years beginning on or after this date.

A standard on how to audit with data is not on the horizon. It is behind us.

Sources: PCAOB Releases 2023-004 and 2024-007; SEC Release No. 34-100774 (20 August 2024).

WHAT THE DRAFT ACTUALLY PROPOSED

Three ideas worth a CA's attention

1

Selecting all items is a named method

The standards set out three ways of selecting items for a test of details: selecting all items, selecting specific items, and audit sampling. Full-population testing is not a novelty. It sits in the standard, next to sampling, as an equal.

2

If the analysis flags it, you must investigate it

Where technology-assisted analysis identifies items within a population for further investigation, the auditor is expected to investigate them. You cannot run the test, see four thousand exceptions, and quietly keep the sixty you like.

3

External data the client hands you is not free evidence

New paragraph .10A of AS 1105 requires the auditor to understand the source and the company's process for receiving, maintaining and processing external information in electronic form — and to test either the information or the relevant controls.

Paraphrased from PCAOB Releases 2023-004 and 2024-007, and the amendments to AS 1105 and AS 2301. Not a substitute for reading the releases.

THE UNCOMFORTABLE COROLLARY

Once you look at the whole population, you own everything you saw.

The fear in the room

“If I run analytics and it throws four thousand exceptions, I have created work I cannot finish and a file I cannot defend. Safer not to look.”

The professional answer

Design the test so the output is risk-ranked and defensible before you run it. Four thousand flags is not a finding — it is a badly designed procedure. Deliberate ignorance is not a defence; poor design is not either.

This is the single biggest cultural barrier in Indian firms today. Name it, then solve it with design.

AND CLOSER TO HOME

The same question, asked in Indian accents

NFRA	2026 inspection reports flag documentation gaps, reliance on management representations without sufficient evidence, and weak external confirmation trails. From FY27, the top six firms face inspections under new guidelines — remediation plans within three months, full implementation within six.
PCAOB inspections	In inspections of 2023 audits, deficiency rates reached 46% of engagements reviewed — cases where the firm had not obtained sufficient appropriate evidence. Reported causes included failure to test the accuracy and completeness of company-provided data, and inappropriate sampling.
The common thread	Every one of these is, underneath, a data question: can you demonstrate that the information supporting your conclusion was complete, accurate and independently obtained?

Sources: NFRA inspection reports and June 2026 inspection guidelines coverage; PCAOB inspection findings as reported by Reuters and the US SEC.

MOVEMENT TWO

Oil is worthless without a refinery

Everybody quotes the slogan. Almost nobody builds the plant.

THE MOST MISQUOTED LINE IN OUR PROFESSION

“Data is the new oil”

What people hear

“Data is valuable. Collect all of it.”

So we buy storage. We buy dashboards. We buy a licence.

And then we wonder why nothing changed in the audit file.

What the metaphor actually says

Crude oil in a barrel earns nothing.

Value is created by cracking, refining and distribution — never by extraction alone.

The world's richest oil firms are refiners and traders, not people who happen to own wet ground.

Most organisations own the oilfield. Very few have built the refinery. That gap is our opportunity.

BUILD THE PLANT

The five stages of an audit data refinery

1 — Extract

Get the population at source — ERP tables and logs, not a management-prepared summary.

2 — Standardise

Clean, type, de-duplicate, reconcile to the trial balance. Prove completeness before you test.

3 — Link

Join master to transaction to log to external source. Value appears at the joins, not in any one table.

4 — Test

Rules, ratios, anomalies, patterns. Full population. Risk-ranked output, never a flat exception dump.

5 — Narrate

Turn exceptions into a decision the client can act on. This is where the CA is irreplaceable.

Stages one and two are seventy per cent of the effort and none of the glamour. Firms that skip them produce confident nonsense.

WHY THE REFINERY NEVER GETS BUILT

Five honest blockers

“The client won't give the data”

Usually means we asked the wrong person for the wrong object. Ask IT for a table, not finance for a report.

“We don't have the tools”

Power Query, Power BI and SQL cover most audit analytics. Tooling is rarely the real constraint.

“No time in the engagement”

True in year one. From year two the same script runs in minutes. It is capital expenditure, not running cost.

“Too many exceptions”

The output was not risk-ranked. An unranked exception list is a failure of design, not of data.

“Our people can't do it”

The scarce skill is not coding. It is knowing which question is worth asking of the data.

FROM DATA TO DECISIONS

The ladder we are actually climbing

Descriptive	What happened?	
Diagnostic	Why did it happen?	*Drill-down, root cause*
Detective	What is unusual?	*Anomaly, Benford, outliers*
Predictive	What is likely next?	*Risk scoring, forecasting*
Decision intelligence	What should we do — and did we do it?	*Recommendation, closed loop*

Most audit 'analytics' stops at rung one and calls it transformation.

AJA Labs.ai

MOVEMENT THREE

Dark data

The evidence your client already owns and nobody has ever looked at

AJA Labs.ai

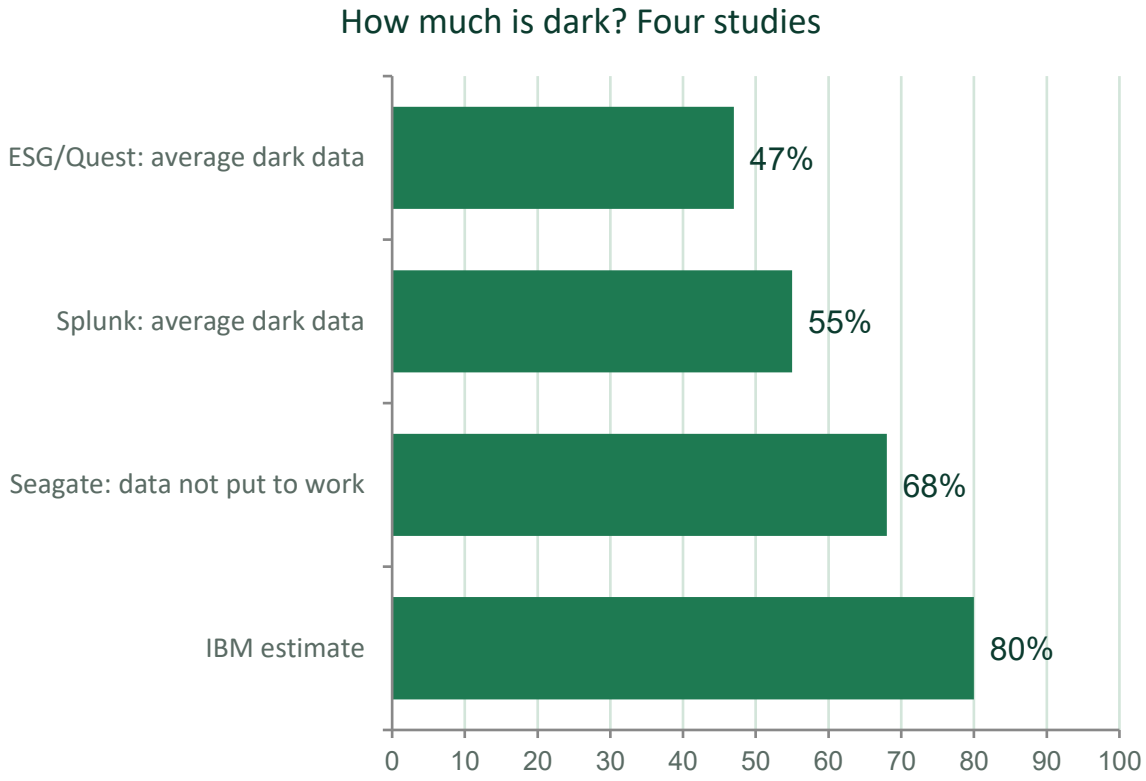
DEFINITION

Dark data: collected, stored, paid for — never used

Gartner's original framing compares it to dark matter: it is most of what exists, it exerts real force, and we cannot see it. Operationally, dark data is any data an organisation generates in the course of business but never uses for analysis or decision-making.

Three reasons it stays dark

- Nobody knows it exists — it was a by-product, not a deliverable.
- It is hard to reach — logs, archives, unstructured text, restricted tables.
- No one owns a question it would answer.



Sources: IBM; Seagate Rethink Data; Splunk dark data survey; Enterprise Strategy Group / Quest, as reported by CIO. Estimates vary with definition — the direction does not.

THE MAP

Dark data hiding inside every ERP you audit

- Master data change logs — who changed a vendor's bank account, and when
- User access and authorisation logs, role assignment history
- Security audit log / SM20 — failed logins, privileged actions
- Workflow approval timestamps and approver substitution
- Document header and item change history (CDHDR / CDPOS)
- Deleted, reversed and parked document trails
- Free-text fields: PO remarks, narrations, credit notes
- Goods receipt versus weighbridge, gate-entry and GPS logs
- Email metadata, attachment hashes, shared-drive access
- Payment returns, rejects and bank statement narrations
- Helpdesk and IT tickets raised around period-end close
- Table logging, transport requests and configuration changes

None of this appears in a trial balance. All of it is admissible audit evidence.

AJA Labs.ai

The books record what was booked.

The logs record what was done.

Fraud, management override and control failure are behaviours. Behaviour leaves its trace in the log, not in the ledger. That is why dark data is not a nice-to-have — it is the only place intent becomes visible.

WHAT IT LOOKS LIKE IN PRACTICE

Three findings the ledger could never give you

The four-second approval

A purchase requisition worth two crore was created and approved within four seconds, at 02:41, by the same user under two roles. Invoice, GRN and payment were all clean. The ledger showed nothing. The workflow log showed everything.

Workflow and user logs

The bank account that travelled

A vendor's bank account was changed two days before payment and restored two days after — three times in one year. Each individual transaction passed three-way match. Only the change history revealed the pattern.

Master data change log

The weight that never matched

Weighbridge tickets averaged 1.8% below GRN quantity for one transporter, consistently, for fourteen months. Within tolerance every single time. Material in aggregate.

Operational log versus ERP

Notice what all three share: every individual transaction was compliant. Only the population revealed the pattern.

Composite illustrations drawn from field engagements; identifying details changed.

WHY IT IS WORTH THE TROUBLE

Four returns on lighting up dark data

1 Evidence quality

Log data is system-generated and independent of the client's preparer. Under SA 500 that is stronger evidence than anything management hands you.

2 Coverage

You move from an opinion supported by sixty items to one supported by the population. The assertion and the test finally match.

3 Early detection

Patterns surface during the year, not after it. Detection time is the single biggest driver of loss size.

4 New revenue

Data readiness, controls monitoring and decision analytics are advisory services your client cannot buy off a shelf.

Dark data is the only audit input where your cost is near zero and your differentiation is near total.

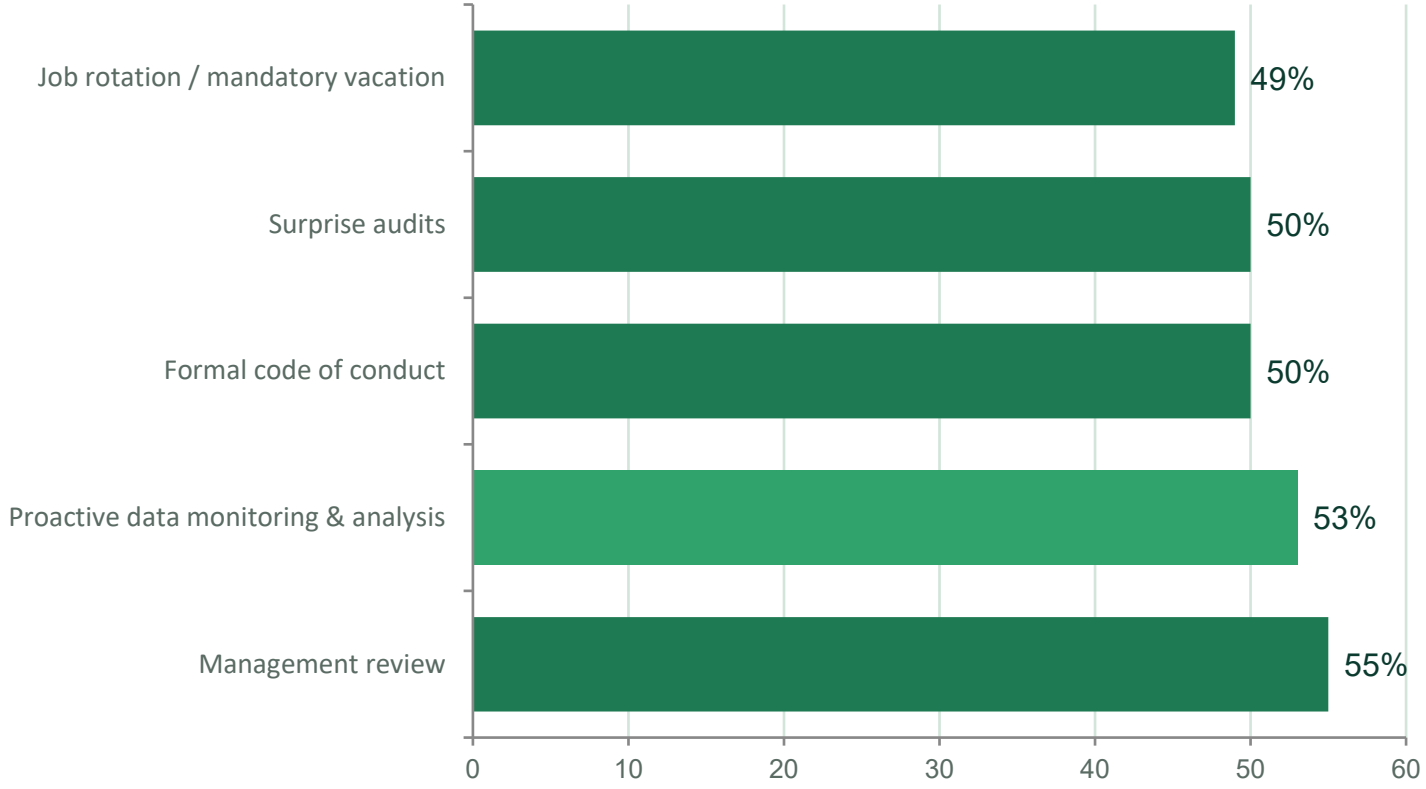
MOVEMENT FOUR

Why data-driven audit wins

Not because it is modern. Because the evidence says so.

THE EVIDENCE — ACFE 2026

Anti-fraud controls, ranked by reduction in median loss



Read this carefully

- Proactive data monitoring is the second most effective control in the world's largest fraud study.
- It sits above surprise audits and above a formal code of conduct.
- And management review, at number one, is only as good as the data placed in front of management.
- Both top controls are, underneath, data problems.

Source: ACFE, Occupational Fraud 2026: A Report to the Nations — 2,402 cases across 143 countries, losses exceeding USD 3.4 billion.

AJA Labs.ai

THE CLOCK IS THE COST

Duration drives loss more than scheme type

5%

of annual revenue lost to occupational fraud by the typical organisation

12

months — the median time a scheme runs before it is detected

$40k

median loss where the scheme is caught within six months

$1.1m+

median loss where the scheme runs beyond five years

An annual audit is, by design, a twelve-month detection lag. Continuous monitoring attacks the one variable that matters most.

Source: ACFE, Occupational Fraud 2026: A Report to the Nations. Figures are medians across the study population.

AJA Labs.ai

SIDE BY SIDE

Sample-based versus data-driven audit

Dimension	Sample-based	Data-driven
Coverage	A judgemental or statistical sample	The full population, every time
What you find	Errors in the items you happened to pick	Patterns and outliers across the whole
Evidence source	Largely management-prepared	Largely system-generated and independent
Repeatability	Re-performed manually each year	A script re-run in minutes; the difference is the finding
Timing	After the year has closed	During the year, as often as you like
Where judgement goes	Into selecting items	Into interpreting anomalies — where it belongs

Data-driven audit does not remove judgement. It moves judgement to the point where it is scarce and valuable.

THIS IS NOT A DEPARTURE FROM THE STANDARDS

The standards already point here

SA 500 — Audit Evidence

Evidence from independent sources is more reliable than evidence generated by the entity. System logs are exactly that. And information produced by the entity must be tested for accuracy and completeness before you rely on it — the same idea PCAOB has now written into AS 1105.10A.

SA 315 and SA 330 — Risk

Understanding the entity includes understanding its information system and the flow of transactions through it. You cannot claim that understanding from a printed report.

SA 240 — Fraud

Requires journal entry testing focused on unusual entries. 'Unusual' is a population concept. It cannot honestly be assessed from a sample.

SA 520 and SA 530

Analytical procedures and sampling both assume you have characterised the population. Data analytics is how you characterise it.

We are not asking permission to do this. We are already required to.

AJA Labs.ai

MOVEMENT FIVE

What you do on Monday

A ladder, six tests and a ninety-day plan

AJA Labs.ai

START HERE

Six tests you can run on your next engagement

1

Vendor and employee overlap

Match vendor master to payroll on bank account, PAN, address, phone. Highest hit rate of any test I run.

2

Bank account change velocity

Flag every vendor bank change within N days of a payment. Then look at who made the change.

3

Postings outside working life

Entries on holidays, after midnight, or by users on approved leave. Cross-check the HR calendar.

4

Approval latency

Time between creation and approval. Anything under a few seconds was not reviewed by a human.

5

Split transactions

Consecutive invoices from one vendor sitting just below an approval threshold.

6

Round sums and near-thresholds

Amounts ending in triple zeros, or clustered at 99% of a limit. Cheap to run, uncomfortable to explain away.

Every one of these runs in Excel Power Query. None of them needs an AI licence.

AJA Labs.ai

LET ME BE HONEST WITH YOU

Five ways this goes wrong

Garbage in, confident garbage out	If the extract is incomplete, every downstream test is worthless — and now it is worthless at scale.
The exception avalanche	Ten thousand unranked exceptions destroy client trust faster than any missed sample. Rank, or do not run.
The black box in the file	If you cannot explain how a model reached a conclusion, it is not audit evidence. Prefer transparent rules to opaque scores.
Privacy and the DPDP Act	Access, workflow and HR data is personal data. Agree scope, purpose and retention in the engagement letter, in writing.
Automation as an alibi	A tool that runs is not a control that works. Someone still has to look, think and sign.

WHERE WE STARTED

If your working paper cannot be re-run, it is not evidence. It is a souvenir.

Three things to take back to your firm

1. **Ask for the table, not the report** — The single highest-leverage change in how you request evidence.
2. **Audit the logs, not only the ledger** — The ledger tells you what was booked. Only the log tells you what was done.
3. **Make one test repeatable this year** — One script, documented, re-runnable. That is your first modern working paper.

Data is not replacing the working paper. Data has become the working paper.

AJA Labs.ai

Ask Me Anything

anand@ajalabs.ai
+91 962 023 3516
www.ajalabs.ai

ajalabs.ai