EFFICIENCY IN DATA RETENTION AND DELETION

Managing Compliance Costs under India's DPDP Framework

Abstract

Boards usually hear about data retention as a compliance checklist item from legal or IT. What it is, is a recurring cost decision. Most organisations make it by default, not by design. Indian companies are sitting on ever-larger piles of customer and employee data: KYC files, call recordings, transaction logs, and behavioural data, backup after backup. Some of it is genuinely needed. A lot of it is simply still there because nobody set a date for it to leave.

The Digital Personal Data Protection Act, 2023, and the notified DPDP Rules, 2025, change that default. Purpose-based retention and erasure move from good practice to governance requirement. Rollout is phased, but the direction is clear. Organisations will need to say why they hold personal data, how long that purpose lasts, whether some other law extends it, and what happens once none of those apply.

This article argues that Cost and Management Accountants are well placed to turn retention from an open-ended obligation into something measurable, proposing a data retention cost curve, a data liability ratio, cost of excess retention, and net compliance efficiency gain. None of this is a case for deleting everything. It's a case for retaining data on purpose: a documented reason, a defined period, and a cost the board has actually seen.

CMA Krishna Prasad Surapaneni
Corporate Professional
Hyderabad
krishnaca.inc@gmail.com

When "just keep it" gets expensive

A pattern shows up repeatedly across data-heavy organisations, NBFCs among them. Cloud spend climbs faster than the business itself is growing. Rarely is there one culprit system behind it. More often it's the slow accumulation of things nobody decided to keep forever: old call recordings, duplicate KYC scans, rejected applications, closed-account files, and backups of backups. None of it was consciously marked "keep indefinitely". It simply never had a deletion date attached in the first place.

Prior to digitalization at scale, archival rooms or storage rooms used to occupy huge areas with files stacking up, costing little overhead beyond rental cost. In the cloud environments, the same files occupy even bigger virtual storage space at scale. Files get indexed, encrypted, monitored, backed up again, and swept into every audit, incident-response, and legal-discovery exercise the organisation runs. A dormant dataset isn't idle. It's still drawing resources and quietly enlarging what a breach would expose.

"Storage is cheap" is true and also beside the point. The storage itself might cost very little. Everything wrapped around it does not: backup, monitoring, log ingestion, access management, encryption, audit testing, legal review, or the eventual effort of deleting it. The storage subscription or access cost is only a smaller part of the real cost stack.

DPDP pushes directly against the instinct to hold

data "just in case". Processing has to stay tied to a lawful, specified purpose. Once consent is withdrawn or that purpose lapses, erasure becomes the default unless some other law says otherwise. For a CMA, that's a plain question worth putting to management: What is the organisation paying, every year, to hold data that no longer serves an active need and isn't legally required to stay?

> Data retention is not only a compliance obligation but a measurable cost decision model requiring lawful purpose, defined timelines, accountable ownership, and informed board-level oversight

Purpose first, cost second

For any dataset of consequence, three questions ought to have clear answers. Why was this collected? Is that reason still alive? Does some other law require it to be kept even after the original reason ends?

The first is where privacy governance meets cost accounting. A genuine purpose should map to a product, a process, or a cost centre. The second identifies the trigger that starts the retention clock: closure, contract completion, an employee's exit, a dispute resolved, or consent withdrawn. The third separates data held by choice from data held because the law says so.

DPDP doesn't sit alone in this. PMLA, RBI directions, the Companies Act, GST law, SEBI frameworks, labour law, contracts, and legal holds may all have something to say about the same dataset. A closed account or withdrawn consent, then, doesn't automatically mean deletion. It might mean moving the data into a locked-down archive, access narrowed, with a clear note explaining the legal basis for keeping it.

The classification that follows is fairly simple in practice. Active-use data stays in live systems. Data kept only because a statute, regulator, or ongoing litigation requires it goes into a restricted archive. Data with no remaining purpose and no remaining legal basis gets erased or properly anonymised. Each category can then be given an owner, a retention period, a storage tier, and a cost.

The one-year rule, and what sits on top of it

The DPDP Rules, 2025, were notified in November 2025 with a phased commencement schedule. Most operational obligations apply only after the transition window closes. That window is the planning period for organisations when retention systems should be redesigned.

Rules 6 and 8 matter most for cost planning. Between them, they call for security logging and, in specified circumstances, retention of personal data, traffic data, and processing logs for at least one year, tied to the purposes listed in the Seventh Schedule. Barring some other law extending it, erasure is expected once the prescribed period lapses or the purpose ends.

A full year of logs adds up fast: login events, failed authentications, API and payment traffic, consent changes, support activity, and breach-investigation material. The cost shows up in SIEM licences, hot and cold storage, indexing, security-operations time, and audit evidence. The real question isn't whether the organisation can retain a year of data. Real consideration is if it can keep them in the right tier, restrict access, monitor usage and actually delete them once the legal basis for retention is over.

Sector rules often push the period further. PMLA generally keeps transaction records for five years, with client identity records carrying their own post-relationship requirements. The Companies Act specified to retaining a minimum of eight financial years for books of account. The CGST Act generally requires seventy-two months from the due date of the relevant annual return, longer where proceedings continue. SEBI-regulated entities carry their own cybersecurity and record-preservation framework on top of all this. There is, in short, no single retention number for an enterprise. Only a reconciliation, dataset by dataset, of DPDP's purpose-based erasure against whatever longer periods other laws demand.

Treating data as both a cost and a liability

It helps to picture personal data moving

through four stages. Call it a Data Retention Cost Curve, for lack of a better name. Both its business value and legal relevance are high during active use. After the purpose ends, data keeps compliance value during the period when statutory requirements still apply but should move to a cheaper and restricted environment. After that comes extended retention with no clear justification, cost accruing while value shrinks. Without this check, it leads to indefinite retention with little or no monitoring and turns into liability from an asset, resulting in risks of exposure. Such liability is something Organizations could not visualise until it hits hard.

> CMAs can convert data retention into a cost-management discipline by identifying excess retention, measuring hidden costs, and evaluating archival, anonymisation, deletion, or automation investments objectively

KYC documents matter during onboarding and the statutory window and become pure liability once that window closes. Call recordings earn their keep for a defined stretch. Kept indefinitely at full fidelity, they're just expensive. Marketing insights data, like Clickstream and abandoned-cart data tends to lose commercial value far faster than transaction records, often within weeks.

To give boards a simple indicator, I propose the following Data Liability Ratio:

Data Liability Ratio = Personal data retained beyond active business or statutory need / Total personal data retained

The ratio will initially depend on sampling, classification and judgement. That does not make it useless. Its purpose is to show whether a material share of the organisation's personal-data footprint is surplus to purpose and to track whether that share is reducing.

A related measure is the Cost of Excess Retention:

Cost of Excess Retention = Data volume beyond approved retention × annual cost per unit

The annual cost per unit should not be limited to storage. It should include a reasonable allocation of backup, security, monitoring, processor, audit and legal-discovery costs. For particularly sensitive datasets, management may also estimate expected risk exposure, although such estimates should be presented separately and with transparent assumptions.

None of this needs an elaborate dashboard. Total personal data held, data under legal hold, data past approved retention, systems with automated deletion, and the annual cost of excess retention. That serves the purpose of turning deletion from IT housekeeping into a governance decision the board actually engages with.

From manual deletion to an automated business case

Organisations tend to move through three stages here. The first is scramble-and-delete: users delete data when they run out of storage space or when an organization-level program for data removal happens. Cheap-looking, but weak on evidence, inconsistent on backups, and prone to missing processor-side deletion or breaching a legal hold. The second is a ticketed workflow: requests logged, routed, reviewed, executed, closed with evidence. More defensible, but it eats hours. Once several functions are spending two or three hours per request, the annual bill adds up fast.

The mature version is rules-based retention and purge. Data classification happens at the stage of collection, is mandated to choose the purpose, is given a legally anchored retention clock, has continuous checks for legal holds and is then deleted or anonymised or archived automatically after the specified period. A working system needs purpose tagging, retention rules, an exception path, deletion logging, processor integration, and reporting that produces evidence rather than just activity.

Worth evaluating like any capital investment. Call it a Net Compliance Efficiency Gain:

NCEG = Avoided storage cost + avoided manual effort + reduced risk exposure + audit efficiency − implementation cost − annual maintenance cost

Take a hypothetical platform processing 500 retention or erasure reviews a month. At 2.5 hours of combined effort per request and a loaded cost of ₹1,200 an hour, that's roughly ₹1.8 crore a year in labour alone. Bring average effort down to 0.6 hours through automation, and the saving is about ₹1.37 crore a year before storage and audit benefits are even counted. The numbers will differ by organisation. The method is what lets management actually test payback rather than buy privacy technology on faith.

A CMA-led retention framework

A CMA-led exercise starts with a proper data inventory and ends with valuating it in rupee terms. Each dataset has to be evaluated for its purpose, ownership, categorization of personal data, contents, volume, retention clock events and costed each month for each of these parameters. These need to be on record.

Each dataset then needs mapping to its actual legal basis rather than guesswork built. They need to undergo an approval process for retention, a periodic review and a clearly defined closure process when the reason no longer applies. The whole repository should be left untouched for lack of these mappings or processes.

Every lifecycle stage carries cost drivers that can be measured. It's often the fragmented, sensitive, multi-processor datasets that turn out expensive, not simply the largest by volume. Management can rank repositories by remaining value, legal necessity, sensitivity, breach impact, and how hard they'd be to delete. Repositories that are costly, sensitive, and no longer valuable go to the top of the remediation list. Where historical data feeds analytics or AI work, the first question should be whether identifiable data is actually necessary at all. Anonymisation often preserves the analytical value while taking the exposure off the table.

Internal audit plays a crucial role in checking that a retention schedule exists, legal holds are reviewed periodically, deletion triggers fire as designed, actual deletion happens as per this process, and most importantly, Board Dashboard carries this status in real time and matches with these actual happenings in the system.

CAs, CSs, and CMAs each bring their own perspective and action plan to this process. CAs on assurance and financial reporting. Company Secretaries on governance and legal documentation. CMAs on quantifying lifecycle cost and flagging avoidable spend. Together, they paint a considerably fuller picture for the Board than any one of these views alone.

Conclusion: deliberate retention, not aggressive deletion

Costs keep climbing and become enormous through holding too much personal data and cloud backup, monitoring, auditing, and discovery. But deleting too early or holding for shorter periods than required leads to risks of not meeting tax and legal requirements. So neither policy of retaining everything nor deleting aggressively is right. The right answer is to retain data deliberately, for a documented purpose and period, in the right system and at a known cost.

CMAs are genuinely well placed to lead this. They can identify the repositories generating the most combined cost and risk, price what it costs to keep each one, test whether a continuing legal or business basis actually exists, and build the case for archival, anonymisation, deletion, or automation, whichever fits.

They can estimate the annual cost of personal data being held beyond the statutory or business purpose right from the start of the budget cycle. Presenting that cost budget along with proposed metrics changes the tenor of the board discussion. DPDP compliance stops beyond the penalties prescribed under the law and consent notices at that point. It becomes part of ordinary cost management and a disciplined answer to one of the fastest-growing invisible liabilities on the modern balance sheet. MA

References

1. *Digital Personal Data Protection Act, 2023*
2. *Digital Personal Data Protection Rules, 2025*
3. *Prevention of Money Laundering Act, 2002, section 12*
4. *Companies Act, 2013, section 128*
5. *Central Goods and Services Tax Act, 2017, section 36*
6. *SEBI Cybersecurity and Cyber Resilience Framework for regulated entities*