Securing the digital lending document workflow

RBI and DPDP obligations at the endpoint, where the loan origination system cannot see

For CISOs, CROs and compliance leaders at Indian NBFCs and lending fintechs | 25 August 2026

Enterprise loan origination systems are audited, penetration-tested and access-controlled. The document that reaches them is not. Between the borrower's e-mail and the LOS upload field sits an underwriter with a 35 MB scanned statement, a 10 MB upload limit, and a search engine.

This paper sets out where that gap sits in the two frameworks an NBFC answers to, what the endpoint failure actually looks like in credit operations, and how the problem changes when document processing happens on the machine rather than in somebody's cloud.

1. The gap the LOS cannot see

An NBFC's document security is usually described in terms of its systems: the loan origination system, the core banking platform, the document management system. All three are audited, access-controlled and, increasingly, penetration-tested.

None of them is where the leak happens. The leak happens in the twenty minutes between a borrower e-mailing a scanned bank statement and that statement reaching the LOS, while an underwriter tries to get a 35 MB file through a 10 MB upload field before the end of the day.

That interval is invisible to every control an NBFC has bought. It leaves no log in the LOS. It appears in no access review. And the tooling used in it is chosen by whoever is doing the work, under time pressure, from the first page of a search engine.

> *This paper describes what software does. It is not legal advice, and no statutory provision cited here has been reviewed by a practitioner on the authors' behalf. Your own adviser should map any obligation to your circumstances.*

2. Two frameworks, one endpoint

2.1 The DPDP Act, 2023

Under the Digital Personal Data Protection Act, 2023, an NBFC handling borrower data is a **Data Fiduciary**: it determines the purpose and the means of processing, and it carries the obligations that follow.

Section 8 requires a Data Fiduciary to implement reasonable security safeguards to prevent a personal data breach. The Act does not enumerate them, which is the point — what is reasonable is judged after an incident, against what the organisation knew and what it did.

When an underwriter uploads a PAN card to a free web converter, that site processes personal data on the NBFC's behalf, without a contract, without a data processing agreement, and generally without the NBFC knowing it happened. Schedule 1 of the Act provides for penalties up to **Rs 250 crore** for failure to take reasonable security safeguards.

2.2 RBI's digital lending framework

The RBI's Master Direction on Digital Lending places responsibility for data handling on the regulated entity, including where the handling is done by a lending service provider or an agent, and constrains where borrower data may be stored. An NBFC does not discharge that responsibility by pointing at a third party it did not choose and cannot name.

A free conversion service routes the file through infrastructure the NBFC has not assessed, in a jurisdiction it has not identified, under terms it has not read. The regulated entity remains answerable for it.

> *Both frameworks are engaged by the same twenty minutes of work, and neither is engaged by a control the NBFC has already bought.*

3. What the failure looks like in credit operations

The scenario is worth stating concretely, because in the abstract it sounds like a training problem and in the specific it does not.

A loan officer receives a six-month bank statement, scanned at 300 DPI, 35 MB. The LOS rejects anything above 10 MB. The officer has eleven files to clear before close. They search for a way to make the file smaller, upload six months of a borrower's transaction history to a service they have never heard of, download an 8 MB file, and push it into the LOS. The file is now in two places, and only one of them is on the NBFC's asset register.

Nothing about this is negligent in the ordinary sense. The officer solved the problem they were given with the tool that presented itself. The control failure is upstream: the organisation created a size limit, provided no compliant way to meet it, and left the choice of tool to the person under the most time pressure.

The three recurring variants

- **Masking that is not masking.** Aadhaar or PAN covered with a highlight, a shape or a black box in a general-purpose PDF viewer. The digits remain in the file, selectable and searchable, and the document now looks handled — which makes it less likely to be checked again.
- **Documents that carry more than they show.** A spreadsheet shared as a working paper still holds its author, its revision history and, often, hidden rows and the workings behind a summary figure. A PDF still names the machine and the software that produced it.
- **Unvetted splitting and merging.** Collateral packs assembled or split with whatever tool was to hand, each pass leaving another copy on another server.

4. Moving the processing to the endpoint

The structural fix is to remove the network from document processing entirely, so that the compliant path is also the fastest one. If the tool on the desktop handles the 35 MB statement in place, nobody searches for an alternative.

Oxofolio is a desktop application for exactly that work. What follows is how it is built, stated so that an infrastructure team can verify each claim rather than accept it.

4.1 Architecture

Three local processes. A main process owns the filesystem. A sandboxed interface process renders documents with context isolation on and no Node integration, so the component that renders untrusted content cannot open a file. A Python sidecar performs the document operations and communicates over newline-delimited JSON on standard input and output.

The application opens no network sockets. Not for licensing, not for updates, not for telemetry. The sidecar speaks over stdio rather than a local HTTP port specifically so this is demonstrable: `netstat` against the running application shows nothing listening and nothing connecting.

CONTROL	HOW IT IS ENFORCED
No network code ships	A scan rejects networking imports across the shipped source and fails the release build. It is not a policy somebody has to remember.
Denied at the OS layer	macOS builds do not request the com.apple.security.network.client entitlement, so the operating system refuses outbound networking whatever the code attempts.
Licensing is offline	Licences are Ed25519 signatures verified against a public key compiled into the application. There is no activation call and no licence server.
No vendor infrastructure	There is no account, no tenant and no vendor-side storage, so there is no environment in which borrower data could sit.

4.2 What it does to a document

OPERATION	WHAT IT DOES, AND HOW IT IS VERIFIED
Compress	Re-encodes the page images, which is where the bytes are in a scan. The saving is shown per file before anything is kept. A page carrying real text is never flattened to pixels to hit a size target.
Unlock in batch	Applies passwords the firm already holds across a folder. Passwords live and die inside the sidecar, are never written to disk, and are asserted absent from every result and log by test.
Find personal data	Aadhaar, PAN, GSTIN and IFSC are checksum-validated rather than pattern-matched, so an invoice total is not mistaken for an identity number. Every match is shown in place before anything is removed.
Redact	The text operators are deleted from the file; the black mark is drawn afterwards as evidence of the removal. The output is then re-opened and searched for the values that were meant to disappear, and reported as failed if any survive.
Strip metadata	Author, producing software, prior titles, revision history and embedded JavaScript can be removed in the same pass.
Record	Every operation writes to a processing log: what was done, when, on which machine, with the SHA-256 of every input and every output.

That last row is the one that matters for a supervisory conversation. It produces contemporaneous, hash-anchored evidence of what was done to which file — the kind of record that is difficult to reconstruct afterwards and straightforward to keep as you go.

5. Where the boundary actually is

A paper that claimed a perfect perimeter would be easy to write and would not survive a competent review. There is one path by which a document leaves the application's control, and it is worth stating precisely because it is the one an assessor will find.

Opening a document in another application on the same machine. Oxofolio can hand a spreadsheet to Excel or Numbers. That opens no socket and the application's own claim survives it intact — but the receiving application may not share its opinions about where client data belongs. Numbers saves into iCloud Drive by default. Excel signed into a Microsoft account turns AutoSave on and places the file in OneDrive.

The application states this at the point of the click rather than in a manual. For an NBFC the mitigation is a policy one: the same MDM configuration that governs OneDrive and iCloud for every other document governs it here, and the handoff can simply not be used on borrower files.

> *This is deliberately not licence-gated or disabled. Preventing a firm from opening its own file in its own spreadsheet application would be a control that trapped data rather than protecting it.*

What the product does not do

- **It does not recover passwords.** There is no brute force, no dictionary and no enumeration. It applies keys the firm already holds. A compliance product that shipped a password cracker would be a liability to the organisation that deployed it.

- **It does not make you compliant.** Compliance is a property of your processes. This produces evidence and does work that would otherwise be done by hand, and no vendor can sell you a legal conclusion.
- **It does not use AI.** Detection and extraction are rules, so the same document yields the same answer next year and a reviewer can argue with the logic.
- **Secure erasure reaches NIST SP 800-88 *Clear*, not Purge or Destroy.** On a solid-state drive, wear levelling can leave the original block intact until it is reused. Full-disk encryption is what makes such residue worthless, and the erasure certificate records whether it was enabled.
- **It does not edit Office files.** The spreadsheet, document and presentation viewers read and never write.

6. Deployment

The rollout is deliberately unremarkable, and the figures below are the published ones rather than a target.

	DETAIL
Platforms	macOS (Intel and Apple silicon), Windows 10 and 11, and Linux as .deb and .rpm.
Installer	Windows is an NSIS .exe installing per-user, so no administrative rights are required. macOS ships as a signed and notarised .dmg.
Size	Between roughly 129 MB and 176 MB depending on platform. The Python runtime, the document libraries and the OCR engine are all bundled, which is why the installer is large and why the application needs no network at any point, including during installation.
Integrity	Every installer is published with its SHA-256 on the download page, so a build can be verified against what was published.
Code signing	macOS is signed with a Developer ID certificate, notarised by Apple and stapled. The Linux .rpm is GPG-signed. The Windows installer is not yet code-signed, so SmartScreen will warn on first run.
Licensing	A signed licence file per machine, verified offline. No activation call, no licence server, and nothing to allow through a firewall.

> *The Windows signing gap is stated here rather than discovered during your evaluation. The certificate is the only outstanding item and the release process fails a build that is unsigned, so it cannot ship by accident while it is being resolved.*

Sequencing

- **Give people the compliant path before you close the other one.** Blocking web converters first produces workarounds; the tool has to be on the desktop when the block lands.
- **Deploy through existing MDM.** Standard installers, no server component, nothing to allow through a firewall.
- **Then restrict the alternatives** at the proxy or CASB, once there is somewhere for the work to go.
- **Verify rather than assume.** Run `netstat` against the application on a managed machine during acceptance testing. It takes ten seconds and it is the whole claim.

7. Conclusion

The endpoint gap in digital lending is not a training failure and it will not close with a policy memo. It exists because an organisation set a constraint, provided no compliant way to meet it, and left the tooling decision to the person with the least time.

Closing it requires that the compliant path also be the convenient one. When the 35 MB statement is handled on the machine in a few seconds, with a hash-stamped record of what was done, nobody has a reason to look for an alternative — and the twenty minutes that no system could see stops being a place where borrower data goes missing.

Every claim in this paper is checkable on a machine you control, which is the only kind of security claim worth making.