The Great Document Migration Nobody Wants to Talk About



Every software purchase comes with a plan for the software. Training sessions. Rollout dates. A kickoff meeting with snacks. Almost none comes with a plan for the files.

That gap is the great document migration. It is the unglamorous part, where twenty years of files move from the old system to the new one. Nobody puts it in the demo. Why would they? "Now watch us move your four million PDFs" is not a crowd-pleaser.

Moving Files Is Harder Than It Sounds

Wikipedia describes data migration as selecting, preparing, extracting, and transforming data before moving it. Notice the first verb. Selecting. Somebody has to decide what moves, and that somebody is rarely volunteering.

The track record is not encouraging. Wikipedia's entry also cites Bloor Research. As of 2011, nearly 40 percent of data migration projects ran over time, ran over budget, or failed outright. That figure is old. Has the underlying problem gotten easier since? Files multiplied, so probably not.

Document migration has an extra wrinkle. Data lives in tidy rows. Documents live in scans, email attachments, and folders named by people who left in 2014.

© All rights reserved | Mindwrap, Inc. 2026 | Terms and Conditions

Most of What You Own Is Junk

Here is the uncomfortable part. HPCwire reported in March 2016 on a Veritas study of more than 2,500 IT executives. More than half of the stored data had no known value. About a third was redundant, obsolete, or trivial. The industry abbreviates that as ROT, which is accurate.

Now run the logic to its extreme. Imagine paying a moving company by the box. Add one box labeled "misc." Inside sits a 2009 lunch menu. Moving it costs exactly what moving your signed contracts costs. Nobody weighs the boxes.

That study is a decade old. Has anyone's file share gotten tidier since? Check yours. The cheapest migration decision is the one made first: what stays behind.

Deleting is not the goal. Retention rules still apply. But a document with no value and no retention duty deserves no seat on the truck.

Folders Are Not Metadata

Think about how your team files things today. A path like Finance\Old\Final\Final2 means something to exactly one person. That person is on vacation. Maybe retired.

Every folder name is a tiny piece of metadata. Client. Year. Status. Document type. Nobody wrote it down as such, because the folder did the job quietly. Move the files without the folders, and the meaning vanishes. Copy the folders without thinking, and you reproduce the mess.

Optix does not organize documents into user-facing folders. Files sit in a system-managed directory under hashed filenames. The database index is how anyone finds a document. So a move into Optix means turning folder habits into index values. Customer. Invoice number. Document type. Date. Once those are set, indexing does the work that Final2 never could.

For large backlogs, Optix includes the Bulk Load Utility, OptixBLU. It reads embedded file metadata, or folder and file names, and maps them into index fields. Jobs save as reusable templates. Every action lands in a log, and runs go unattended. Not every backlog fits those two patterns. Other programmatic routes exist, and the right one depends on where the files come from and what they look like.

The Paper Is Still There

Folders are only half the problem. Somewhere in a basement, boxes of paper are waiting for their turn. Some were scanned in 2011 by a vendor who named files by date. Others were never scanned.

Your basement is not unusual. AIIM, the industry association, surveyed 600 enterprises in 2025. The survey found that 61 percent of intelligent document processing processes still include paper.

Scans cause a quieter problem. A scanned page is a picture of text, not text. Until OCR reads it, nothing can search it, classify it, or route it. We made the same point about search in Semantic Search and Other Things That Should Have Existed Years Ago. Our guide to scanning in-house or outsourcing covers the paper side.

© All rights reserved | Mindwrap, Inc. 2026 | Terms and Conditions

Migration is the best moment to fix this. Optix captures documents and runs OCR on scans. A large language model, OpenAI or Gemini, then reads the text and identifies the document type. It is a general-purpose model, not one trained on your files. On standard business documents, it performs well in Mindwrap's testing. Check a sample anyway. That is step five.

Retention and Regulation Follow the Files

Moving a document does not cancel its obligations. Retention periods, legal holds, and audit trails travel with it. Or they should.

Federal rules make this explicit. 36 CFR 1236.14 applies to U.S. federal agencies. If a record's retention outlasts its system, they must plan the migration before retiring that system. They must also keep each record linked to its metadata. And they must capture the relevant metadata at the point of migration. Private companies are not bound by that rule. It still makes a decent checklist.

There is also a standard for the job itself. ISO 13008:2022 covers digital records conversion and migration. It replaced the 2012 edition. ARMA Magazine described how the Calgary Police Service moved more than 3.9 million case files. Its legacy system had run for over 40 years. The team applied the 2012 edition. An independent reviewer compared random samples by hand. Cutover happened over a weekend. If police evidence can survive that process, your invoices can too.

Compliance teams care about all of this. Our post on the compliance people are hiring covers the retention side.

How to Run It Without Losing Your Mind

A migration is a project, not a weekend. Still, the sensible version has five steps.

1. Decide what stays behind. Check retention rules first, then delete or archive the rest.
2. Map every folder habit to an index field. Write the mapping down.
3. Pilot with your worst documents. Crooked scans. Misnamed files. Anything labeled "misc."
4. Reconcile counts. Files out of the old system should equal files into the new one.
5. Have someone independent check random samples by hand.

Keep the old system read-only until sign-off. Pulling the plug on day one is bold, and boldness is not a validation method.

The Bottom Line

Document migration is the part of every system change that nobody budgets for. It is also where the new system either starts clean or inherits twenty years of mess.

Buying the software is the easy step. Deciding what to move, what to translate, and what to leave is the real work. Skip that, and the new system holds the old chaos at a higher resolution. Is that an upgrade? Technically.

Contact us today to see how Optix by Mindwrap turns folder habits into searchable index values. Then workflow rules route each document onward. Your migration starts with a plan for the files.

© All rights reserved | Mindwrap, Inc. 2026 | Terms and Conditions