ADLIB

ADLIB + VEEVA

Fueling Smarter Submissions with Veeva

What Defensible AI Actually Requires in a Validated Life Sciences Environment

Veeva Vault is the system of record your regulated content lives in. This guide explains what sits beneath it, why that layer is now part of the regulatory conversation, what your validation obligations look like either way, and how Adlib fits into a Veeva environment.

2026 Edition | Adlib Software

Contents

1. How These Two Layers Fit Together

Most life sciences organizations running Veeva believe they have already solved for defensible AI. They have a validated system of record. Documents go in, get versioned, get approved, get audited. Vault even converts files to PDF automatically. From where most teams sit, that looks like the whole job.

It is a reasonable read, and it is incomplete, not because Vault is doing anything wrong, but because “defensible AI” and “a validated document management system” are answers to two different questions. Vault answers: is this the correct, approved, current version of this document? AI answers a different question: can I prove what this output is based on, down to the page, the version, and the moment it was read? Veeva was built to do the first job extremely well. It was not built to do the second, and it does not claim to be.

This guide fills in the rest of the picture: what breaks in a regulatory operation today with no AI involved at all, what regulators are asking for as AI use accelerates, what Veeva provides, where the gap sits, who carries the validation burden, and the objections your own organization will raise when you take this internally.

> *Adlib is a data connectivity solution that sits between the end users and the final destination for their documents, often Veeva or another regulatory information management system. We automate the steps required to prepare those documents.*
>
> **Anthony Vigliotti, Chief Product Officer, Adlib**

BOTTOM LINE

Veeva governs the record. Adlib governs what happens to a document, and to what AI reads from it, before and after it passes through that record. Both jobs matter more than they did a year ago, because the thing reading your documents next is often an agent, not a person.

2. What This Is Costing You Today: Before AI Enters the Picture

If you strip every mention of AI out of this guide, the document layer is still a problem. It is worth starting there, because the cost is already on your books this quarter.

THE FOUR PLACES IT SHOWS UP

- **The gap between last patient last visit and filing.** Regulatory teams routinely lose weeks of that window to formatting, publishing, and reconciliation work that has nothing to do with the science. Every week is a week of patent life and market exclusivity you do not get back.
- **Technical deficiencies, not scientific ones.** Submissions get held up over hyperlinking, bookmarking, PDF/A conformance, file naming, and cross-module inconsistency: errors made by subject matter experts who were never trained on eCTD technical rules and should not have to be.
- **Content arriving from outside your walls.** CRO deliverables, partner submissions, manufacturing site records, and files inherited through acquisition all show up in whatever shape the sender chose. Someone on your team normalizes them by hand.
- **The publishing bottleneck.** A small number of people who know how to assemble a compliant sequence become the constraint on everything, and they are the hardest role in the function to hire for.

None of that requires a model, an agent, or a pilot to fix. It requires the document layer underneath your system of record to do more than store and display.

> **BOTTOM LINE**
> The AI argument in this guide is a reason the document layer is about to matter more. It is not the reason it matters. If you never deploy a single agent, the formatting, publishing, and inbound-content problems are still yours.

3. The State of AI in Life Sciences, in Numbers

AI adoption in life sciences is real, it is accelerating, and it is running well ahead of the governance most organizations have in place to support it.

At the Clinical Trials Technology Congress in London in May 2026, the Pistoia Alliance polled clinical trial professionals on what was holding AI back. Half named trust and regulatory uncertainty as the single biggest barrier, ahead of cost, ahead of talent, ahead of the model itself. In the same poll, roughly 42 percent said they had already seen early ROI from AI.

Read those two findings together and the picture is clear. Organizations are not struggling to find a use case. They are struggling to make the use case defensible enough to scale past a pilot.

A separate 2025 vendor survey of life sciences leaders found that fewer than half were fully confident their data was accurate, timely, and consistently available, while nearly all agreed trusted data was essential to their AI strategy. Treat vendor-published survey figures as directional rather than authoritative; the gap between those two numbers is the useful part, not the precision of either one.

> **BOTTOM LINE**
> The industry does not have an enthusiasm problem or a use-case problem. It has a proof problem: leaders can show AI works in a demo, and struggle to show, twelve months later, exactly what it was based on.

4. What Regulators Actually Require, and What an Inspector Asks For

You already know the guidance exists. The more useful question is what it turns into when someone is sitting in your conference room asking questions.

THE GUIDANCE LANDSCAPE, BRIEFLY

- FDA draft guidance, “Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products” (January 2025, docket FDA-2024-D-4689). Still in draft.
- FDA and EMA joint “Guiding Principles of Good AI Practice in Drug Development” (January 14, 2026). Ten principles, and the strongest cross-agency signal to date.
- FDA guidance for AI-enabled device software functions (final, August 2025), which establishes predetermined change control plans for AI/ML components.
- The CDER AI Council, established 2024, coordinating AI oversight and policy across the center.

Three of the ten joint principles speak to documents and data rather than model performance: context of use, data governance (provenance and processing steps documented in a way that is traceable and verifiable in line with GxP), and lifecycle management. None of them prohibit AI in regulated workflows. They shift the burden from asking permission to providing proof.

WHAT THAT LOOKS LIKE ACROSS THE TABLE

Here is the version that matters. An inspector or auditor picks a statement in a submission you filed fourteen months ago and asks four questions in sequence.

THEY ASK	YOU NEED TO PRODUCE
Where did this figure come from?	The specific source document, the specific page, and the specific version that was current at the time, not the version that is current now.
How was it extracted from that document?	The processing step, the date it ran, and the confidence associated with that extraction. “Our system reads PDFs” is not an answer.
Who or what reviewed it, and when?	A review record tied to the extraction, with author and timestamp, showing the human judgment applied and the disposition reached.
The source document was superseded in March. What happened to everything derived from it?	Evidence that the derived content was retired, re-processed, or flagged. This is the question most organizations cannot currently answer.

Question four is where the room usually goes quiet. Your Vault is validated and your document lifecycle is governed. The extracted text, the chunks, and the embeddings your AI actually queries typically are not, and they usually do not get retired when the source does.

> **BOTTOM LINE**
> None of the current guidance asks which model you used. All of it asks some version of: can you prove where this came from, and can you still prove it a year from now?

5. What Veeva Does Extremely Well

Veeva Vault earned its position as the system of record for regulated life sciences content because it does its core job well: version control, approval workflows, access permissions, and audit trails that hold up to inspection. Any conversation about AI governance in a Veeva environment should start from that fact, not around it.

Vault also includes a genuinely useful built-in capability that is easy to underestimate. It automatically generates a viewable PDF rendition of nearly any file a user uploads (Microsoft 365 files, images, HTML, email formats), so documents can be viewed and searched inside Vault without a separate conversion step. For scanned documents it runs OCR, and it preserves useful structure like bookmarks and clickable links. For what it is designed to do, it works well and it works out of the box.

WHAT THAT FEATURE IS BUILT FOR

The rendition feature is a viewing and searchability tool, and Veeva is explicit about its scope. OCR text extraction caps out at 1,500 pages. Password-protected Microsoft 365 files cannot generate a rendition at all. Very large or low-confidence scans can fail to process. PowerPoint speaker notes are not carried into the rendition, and some secured PDFs cannot render a viewable copy. None of that is a flaw. It is the natural scope of a feature built to make an approved document easy to open and read.

> **BOTTOM LINE**
> Vault’s rendition engine solves “can a person open and search this document inside Vault?” and solves it well. It was never scoped to solve “can an AI system trust this document’s structure, and can we prove that six months from now?” That is a different requirement.

6. The Gap Between “Viewable” and “AI-Evidentiary”

A viewable, searchable PDF is enough for a human reviewer. It is usually not enough for an AI system that has to produce an output someone can trace back to its source under audit conditions.

WHAT AI-EVIDENTIARY REQUIRES THAT VIEWABLE DOES NOT

- **Structural preservation, not just readability.** A model needs headers, tables, and hierarchy intact and machine-readable, not merely visible to a human eye. A rendition can look correct on screen while the underlying structure is flattened.
- **Confidence scoring on extraction.** Not just that text was extracted, but how reliable that extraction was, page by page, so low-confidence content gets routed for review instead of fed to a model as if it were certain.
- **Format coverage beyond the common case.** CAD drawings, proprietary lab and manufacturing formats, and the long tail of legacy file types that arrive from CROs, partners, and manufacturing sites.
- **Cross-document and completeness validation.** Checking that a submission package, TMF artifact, or correspondence bundle is complete against a defined standard, not just that each file individually renders.
- **Provenance that survives past the project team.** A record of which version of which source document, ingested through which channel, on which date, that a compliance reviewer can reconstruct with nobody from the original team in the room.

A document management system tracks the document. AI reads a derivative of it: extracted text, chunks, embeddings. That derivative layer sits outside the validated boundary that governs the source document. Unless something is specifically built to govern it, it is unversioned and it does not get retired when the source document does.

> **BOTTOM LINE**
> If an AI output cannot point to the specific page, version, and extraction confidence it came from, it is not defensible, regardless of how well the source document rendered inside Vault.

7. Intake, Transform, Data: Where Adlib Fits

Adlib's role in a Veeva environment breaks into three connected jobs. Naming them separately matters, because each closes a different part of the gap.

Intake: capturing content wherever it actually originates

Documents arrive from everywhere: inboxes with mixed attachments, partner portals, scanned batches from CROs, file shares. Adlib captures documents from these sources and immediately begins standardizing them: file-type detection, scan-quality cleanup, and an initial pass at metadata, before anything reaches a Vault workflow.

Transform: turning raw input into AI-ready, Vault-ready output

This is the core of what Adlib has always done, and it goes well beyond rendition. OCR and de-skewing across scanned pages. Format conversion across 300-plus file types including CAD and legacy formats. Merging and reordering. Tables of contents and bookmarks. Watermarks and digital signatures. Validation against PDF/A and other technical submission standards. The result preserves structure, which is what separates a document a model can search from one it can reason over.

Data: moving from processed documents to a governed record

The next problem past clean documents in Veeva is this: once an agent extracts something from a document (a clause, a data point, a commitment), where does that extracted fact live, and how do you know which version of the source it came from when the source is later superseded? Adlib's direction is a governed record of what was extracted, from which version, on which date, with what confidence.

> **BOTTOM LINE**
> Intake gets the document in the door in a shape Vault can use. Transform makes it structurally usable. The data layer makes what was extracted from it provable, months later, to someone who was not there.

8. Two Roadmaps, One Direction: Vault AI, Falcon, and Adlib

Veeva has moved quickly on AI, and it is worth engaging with what they have actually built. At its 2026 Summit, Veeva introduced Vault AI: agents embedded directly in Vault for document chat, translation, summarization, and comparison, operating on data already inside a Vault instance under existing permissions. Veeva also introduced Falcon, an agentic platform delivering standing agents for high-volume repetitive workflows, including TMF document intake and quality checks, safety case intake, and health authority interaction management.

Both are useful additions, and both target a real problem. Falcon's clinical agents, based on Veeva's public materials, work natively within Vault applications, using the data structures, workflows, and compliance logic already built into the ecosystem. That is exactly where an agent operating on validated, in-system content should live.

Adlib sits upstream. Vault AI and Falcon are strongest when what reaches them is already clean, structured, and complete. A scanned consent form, a CAD drawing, a fifteen-year-old contract archive, or an inbox of mixed attachments is not yet in a shape those agents are optimized to act on.

THE FAIR QUESTION: WON'T VEEVA JUST BUILD THIS?

You should ask it. Here is the honest answer.

Veeva builds for content inside Veeva. That is not a limitation, it is the product strategy, and it is the right one for a system of record. But a large share of the content that determines whether your submission is defensible never sits in Vault, or does not sit there yet: an FDA information request that lands as an email attachment, CRO deliverables staged on a partner file share, CMC records in a manufacturing system, a Documentum archive nobody migrated, a decade of content inherited through an acquisition. Veeva has no commercial reason to solve for content outside its own boundary, and a vendor whose value depends on being system-agnostic has every reason to.

The narrower version of the same question (will Veeva extend rendition to cover CAD and proprietary lab formats) is possible. It has not happened across many years of the platform's life, and format coverage is a long-tail engineering problem with no strategic payoff for a system-of-record vendor. Weigh that how you like. It is your risk assessment, not ours.

> **BOTTOM LINE**
> Vault AI and Falcon are Veeva building smart agents on top of Vault-native data. Adlib decides what is allowed to reach that data, in what shape, from systems Veeva does not own. The two roadmaps point the same direction and strengthen each other.

9. How It Works: Before, Inside, and After Veeva

Before documents enter Veeva: intake and transformation

Adlib takes documents from email, SharePoint, file shares, and scanned forms and processes them automatically: file-type detection, scan cleanup, key data extraction, metadata tagging, and conversion into standard compliant formats. Consistency regardless of where a document came from or what condition it arrived in.

While documents live in Veeva: ongoing enhancement

Adlib continues adding value after documents are stored: filling in missing metadata, resolving formatting issues, and keeping documents aligned to Vault-specific requirements like submission formatting and audit trail standards as those requirements evolve.

When documents leave Veeva: assembly for submission or for agents

When it is time to build a submission, an audit package, or a grounded input for an AI agent, Adlib handles the assembly: combining documents into packages, applying submission-specific formatting, validating against regulatory requirements, and preparing content for downstream processing, human or automated.

THREE WAYS ADLIB CONNECTS TO WHAT YOU ALREADY RUN

- Out-of-the-box connectors for Veeva Vault, SharePoint, ECM and RIM platforms, CTMS, eTMF systems, ERP, and data lakes; no custom integration work required.
- A REST API for teams building custom pipelines: submit documents, receive structured outputs (PDF/A, JSON data contracts, confidence scores, provenance metadata).
- MCP (Model Context Protocol) support, so AI agents (Claude, ChatGPT, Copilot, and others, Veeva's own agents included) can call Adlib directly at inference time and use validated, citation-anchored content as grounded evidence.

10. Validation and Compliance: What Adlib Carries, What You Own

This is the section your quality organization will turn to first, so it goes before the business case rather than in an appendix.

Adding any component to a validated environment creates a validation obligation. The question is not whether one exists; it is how much of it lands on your team, and how much the vendor absorbs. Here is the split.

AREA	WHAT ADLIB PROVIDES	WHAT YOUR TEAM OWNS
Qualification documentation	Vendor-supplied IQ/OQ documentation and test evidence package.	PQ against your own intended use, and the decision on CSV versus a CSA-aligned risk-based approach.
21 CFR Part 11 / EU Annex 11	Audit trail, access control, and electronic signature controls at the platform layer.	Procedural controls, SOP alignment, and periodic review of the audit trail.
GAMP 5 categorization	Guidance on typical categorization and supporting documentation.	Final categorization decision, which depends on how you configure it.
Change and revalidation	Release notes, change notification, and regression evidence for platform updates.	Impact assessment and any revalidation triggered by your own configuration changes.
Vendor qualification	Completed vendor assessment questionnaire, quality manual, and audit rights under contract.	Supplier qualification decision and periodic re-assessment per your SOP.
Data residency and hosting	Deployment options and documented processing locations.	Privacy impact assessment and any transfer mechanism required under GDPR.

One point worth stating plainly, once the specifics above are confirmed: **the validation burden for the document processing layer is a burden somebody carries permanently.** A vendor absorbs it as a cost of doing business across its whole customer base. Built internally, it becomes a standing obligation for a team that has other priorities, which is the subject of the next section.

11. The Four Objections You Will Hear Internally

If you are the person bringing this forward, you will not be arguing with Adlib. You will be arguing with your own IT organization, your quality group, and whoever controls the budget. These are the four you should be ready for.

“We already run Veeva Vault.”

Vault manages where content lives and proves the record is correct. It does not parse an inbound health authority query, trace a claim back to its CTD source section, validate consistency across modules, or govern the derivative content your AI actually reads. Those sit upstream and downstream of the record. This is an addition to a Vault investment, not an alternative to one.

“We already have OCR and IDP tools.”

Extraction and regulatory-grade traceability are different problems. Most IDP tooling tells you what it read. It does not tell you how confident it was on page 340, which version of the source it read, whether the package was complete against a defined standard, or what happened to the extraction when the source document was superseded. That last one is the audit question, and it is the one general-purpose tooling does not answer.

“We use Copilot and ChatGPT for this already.”

This is not a competing position. A general model cannot point to Module 2.7.4, Section 2.8.1 and show you the paragraph it drew from. It produces fluent output that reads well and cannot be defended. The point of a governed evidence layer is to make the general-purpose model safe to use in a regulated workflow, not to replace it.

“Our Digital R&D team could build this.”

They probably could build the easy part. Document conversion, OCR, and a retrieval pipeline over a cloud extraction service is a credible six-month project, and it will handle a large share of your volume well.

The three things that make the build harder than the estimate:

- **The long tail is where the time goes.** The last stretch of format coverage (CAD, proprietary lab and manufacturing formats, corrupted and secured legacy PDFs, mixed-attachment email) consumes more engineering effort than the first 70 percent of volume, and it is exactly the content that shows up in submissions.
- **ALCOA+ and CTD hierarchy preservation are not a one-time build.** They are a continuing validation obligation. Every model update, library version, and new source format is a revalidation event. Somebody internal now owns that forever, alongside their other work.
- **The liability transfers with the build.** When an internal tool drifts and an extraction turns out to be wrong in a filed submission, the audit trail leads to your own team. With a qualified vendor, the vendor carries qualification evidence, change control, and audit support as a contractual obligation.

The right way to frame this to a CIO is not build versus buy on features. It is where you want the permanent compliance obligation to sit.

> **BOTTOM LINE**
> Every one of these objections is reasonable. Three of them are resolved by scope: different layer, different problem. The fourth is a real decision, and it should be made on who carries the ongoing validation burden, not on whether the first version is technically achievable.

12. The Business Case

Both the FDA and EMA have signaled a longer-term direction toward structured data packages rather than traditional document submissions. Neither has published a firm date, but the direction is consistent enough across their public materials to plan around.

> *Think about what happens today when the FDA receives a submission. Someone has to open it, read it, comprehend it, and sort through thousands upon thousands of pages to make sense of it. Now imagine a future where, instead of receiving documents, the FDA directly receives the data that matters for the submission. The document becomes a backup, an archive. The data is what drives decision-making. That is the vision.*
>
> **Kristen Sauter, President & GM of Life Sciences, Adlib**

Organizations that build clean, governed data pipelines now will be better positioned when that shift formalizes, not because the timeline is known, but because the underlying work does not disappear regardless of when regulators finalize the requirement.

BUILD THE CASE WITH YOUR OWN NUMBERS

Four inputs, all of which you already know or can get in an afternoon. Fill them in before you take this to a steering committee; a general argument about efficiency will not survive that room.

INPUT	WHERE TO GET IT	YOUR NUMBER
Regulatory submissions and major amendments per year	Regulatory operations	
Hours per submission spent on formatting, publishing, and reconciliation	Publishing team: ask for actual, not planned	
Share of deficiencies and information requests that are technical rather than scientific	Your last twelve months of health authority correspondence	
Revenue per month for the lead product in the affected pipeline	Commercial or finance	

The fourth input usually settles the argument on its own. A product doing 400 million dollars a year earns roughly 33 million dollars a month. Against that, a two-week acceleration in a filing timeline is worth more than the entire document infrastructure line item, and nobody in the room needs a spreadsheet to see it.

The first three inputs give you the recurring, defensible number, the one that holds up when someone points out that not every delay is on the critical path.

PROOF AT SCALE

Adlib is deployed in production inside validated, Part 11-governed environments across a substantial share of the largest pharmaceutical manufacturers, including one top-ten pharma processing well over 100 million pages annually. This is core document transformation and rendering running live at enterprise scale, not a pilot, and not a reference architecture.

> **BOTTOM LINE**
>
> Veeva governs the broader process of bringing a product to market. Adlib keeps the data feeding that process clean and provable, whether it lives in Vault eTMF, Vault QMS, or a system outside the Veeva ecosystem entirely.

13. Seven Use Cases

The first four sit squarely inside a Veeva workflow. The last three involve content that starts, or stays, outside it, which is where most of the unaddressed cost tends to be.

USE CASE	THE CHALLENGE	HOW ADLIB HELPS
Automating external document intake	Every employee receiving an email with attachments repeats the same manual process: save, convert, name, and file correctly in Veeva, multiplied across tens of thousands of documents in a single trial.	Incoming email routes to an Adlib-connected inbox. Adlib extracts metadata and attachments, converts diverse file types to standardized PDFs, applies OCR and cleanup, classifies content, and delivers it into the correct Vault location.
Technical compliance for regulatory submissions	Subject matter experts who create clinical, quality, and manufacturing content are not trained on FDA and EMA technical formatting rules. A technical error, not a scientific one, is a common reason submissions are held up.	Adlib applies formatting, validation, and compliance checks automatically so experts focus on content while every document meets technical requirements before it reaches a Vault workflow.
Extending format coverage beyond Office-centric files	Vault's rendition engine is built around the file types most teams use daily. QA, manufacturing, and research teams also work in CAD, proprietary lab formats, and dozens of others outside that scope.	Adlib supports 300-plus file types, converting and standardizing anything that does not arrive Office-ready so it can flow into Veeva without a blind spot.
Preparing grounded content for AI agents and RAG	Models and agents given raw, unprepared documents produce inconsistent results: hallucinations rise, retrieval pulls the wrong chunk, and outputs are not reproducible.	Adlib provides structured, citation-anchored, confidence-scored content so agents, Veeva's own included, and RAG pipelines work from trusted input rather than ungoverned text.
Legacy migration and archive remediation	Content inherited through acquisition, or left behind in Documentum and SharePoint during a Vault migration, is inconsistent, unclassified, and frequently unsearchable, and it is still subject to retention and inspection.	Adlib processes archives at volume: normalizing formats, applying OCR, extracting and mapping metadata, and validating output so historical content lands in Vault usable rather than merely stored.
CRO and partner content normalization	Outsourced clinical execution means the documents that cause the most pain arrive from someone else's system in someone else's conventions. Your team normalizes them by hand, under timeline pressure.	Adlib applies a single standard to inbound partner content at the point of receipt, regardless of source system or file type, so quality does not depend on which CRO sent it.
Health authority correspondence intake and response support*	An information request arrives as an email attachment. Someone reads it, works out which CTD sections are implicated, finds the right SME, and assembles context by hand, while a response clock runs.	Adlib parses inbound correspondence, extracts deadlines and commitments, identifies and retrieves the implicated source sections, and assembles a routing package with citations intact so the SME starts from evidence rather than a blank page.

* In development

14. Next Steps: Three Questions to Ask This Week

Document infrastructure rarely comes out of a regulatory affairs budget. It usually sits with IT or R&D IT, which means your first move is internal, not external. Three questions, in this order.

1. ASK IT WHETHER ADLIB IS ALREADY DEPLOYED HERE

This is not a rhetorical question. Adlib is embedded in a large number of enterprise pharma environments as a rendering and transformation layer, often procured years ago by IT and invisible to the regulatory function. If the answer is yes, this stops being a procurement conversation and becomes an extension of something already validated and already approved.

2. ASK YOUR PUBLISHING TEAM WHERE THE HOURS ACTUALLY GO

Not the estimate. The actual breakdown on the last three submissions, split between scientific content, technical formatting, and rework. That number is your business case, and you cannot get it from a vendor.

3. ASK WHOEVER OWNS YOUR AI GOVERNANCE COMMITTEE QUESTION FOUR

From Section 4: when a source document is superseded, what happens to everything derived from it? If nobody owns the answer, you have found the gap this guide is about, and you have found it before an inspector did.

Ready to see how this works with your Veeva implementation?

Schedule a working session with Adlib's life sciences team.

Bring your publishing hours and your last twelve months of technical deficiencies; the conversation is better with your numbers than with ours.

Schedule a working session ›

About Adlib

Adlib is a document transformation and data connectivity platform for regulated industries. It sits between the systems where content originates and the system of record where it must land, automating intake, classification, rendering, and validation so teams and AI agents work from consistent, citation-anchored, audit-ready output.

Adlib runs in production inside validated, Part 11-governed environments at many of the largest pharmaceutical manufacturers, with connectors for Veeva Vault, SharePoint, ECM and RIM platforms, CTMS, eTMF, ERP, and data lakes.

Learn more at **adlibsoftware.com**