MPL LEGAL TECH ADVISORS

The Legal Data Foundation Blueprint

Most law firms think they have “data,” but not structured, verified evidence that AI systems can actually use.

Broken folder hierarchies, missing metadata, and scattered drives make contract review automation impossible.

This blueprint distills the same foundation we build in client pilots. A step-by-step structure your team can implement safely, before you deploy any AI.

1.) Map the real workflow

Objective: Capture the actual contract intake and triage flow.

Step by step:

1. Shadow one real incoming contract from receipt to filing.
2. Ask the record:
    a. Where did it arrive?
    b. Who touched it first?
    c. How did they determine the document type?
    d. Where was it saved?
    e. Who reviewed or escalated it?
3. Collect the artifacts: intake email, attachments, filenames, and escalation notes. Store them in `/Evidence/WorkflowStudy`.
4. Map the flow on a single page with six lanes:
    a. Client Actions
    b. Frontstage
    c. Backstage
    d. Tech
    e. Evidence
    f. Acceptance

Book a 15min AI & Data Readiness Check

2.) Build the clean repository

Objective: Stand up a clean, matter-centric document library. Never retrofit legacy drives.

Folder structure:

```
Legal-Contracts/
├── Clients/
│   ├── [ClientName]/
│   │   ├── [MatterCode]/
│   │   │   ├── Intake/
│   │   │   ├── Negotiation/
│   │   │   ├── Signed/
│   │   │   ├── Amendments/
│   │   │   └── _Evidence/
```

Naming convention:
{MatterCode}_{DocType}_{Counterparty}_{YYYYMMDD}_v01

Example:
MSA-2025-014_MSA_AcmeCo_20251001_v01.docx

Required metadata fields:

Field	Purpose
Counterparty	Enables search and filtering
EffectiveDate	Time-based reporting
RenewalDate	Contract lifecycle automation
GoverningLaw	Jurisdictional triage
LiabilityCap	Risk benchmarking
PlaybookVersion	Routing reference
Deviation %	Scoring input
Triage Level	Workflow routing

Additional context fields (per matter):

MatterID, Client, Jurisdiction, Status

Book a 15min AI & Data Readiness Check

3.) Pilot one document type (NDAs)

Objective:
Prove repository structure and metadata consistency with one document type.

How to read between the lines of vendor privacy policies: AI vendors often sugar-coat their data usage terms. Here's how to decode some common phrases you'll encounter:

Step by step:

1. **Select pilot set** (20–50 NDAs from past 6–12 months)
2. **Rename** (apply standard naming convention)
3. **Tag metadata** (Complete Counterparty, EffectiveDate, RenewalDate, GoverningLaw, LiabilityCap, PlaybookVersion)
4. **Define extraction rules:**
    a. *EffectiveDate* = first date in preamble
    b. *Renewal* = clause containing "auto-renew"
    c. *GoverningLaw* = heading "Governing Law"
    d. *LiabilityCap* = numeric + currency near "liability shall not exceed"
5. **Handle scan:**
    a. OCR at 300–400 DPI. Spot-check 10–20%
    b. Unreadable files → `ManualReview`
6. **Evidence** (Save extractor JSON + decisions in `/_Evidence`)

Metric	Target	Action
Rename accuracy	≥ 98%	Rework if lower
Metadata accuracy	≥ 95%	Correct within 48h
Critical errors	≤ 1%	Root-cause + re-run
Evidence completeness	100%	Mandatory

Book a 15min AI & Data Readiness Check

4.) Extraction prompt pack

Objective:
Extract structured, verifiable data for downstream triage.

Field	Source Rule
EffectiveDate	First date in preamble
Renewal	Clause with “auto-renew”
GoverningLaw	“Governing Law” / “Applicable Law”
LiabilityCap	“shall not exceed” phrase
Termination	Clause containing “terminate”
Data	Clause containing “data”, “privacy”, or “security”

- Write extracted values back to SharePoint columns.
- Save full JSON under `/Evidence/[Matter]/ExtractorOutput.`
- Include `document_hash and timestamp.`

Quality assurance:

1. 100% manual review on first run.
2. Then 20% sample + all fails.
3. If same error repeats 3×, update prompt pack.
4. Re-run only the affected files.

Book a 15min AI & Data Readiness Check

5.) Convert extracted data into playbooks

Objective:

Transform structured data into defensible routing logic.

Step by step:

1. Use signed MSAs from past 12-24 months
2. Export extractor JSON and CSV
3. Keep latest signed per matter, drop drafts.

Pivot liability caps:

- ≤12 months = Preferred
- ≤18 months = Acceptable
- 18 months = Escalate
- Uncapped = Red flag

Calculate deviation percent:

(Out-of-bounds clauses ÷ Total checked) × 100

Routing logic:

- <20% → Paralegal
- 20–60% → Associate
- 60% → Senior

Book a 15min AI & Data Readiness Check

6.) Intake triage by deviation

Objective:

Replace subjective judgment with measurable deviation scoring.

Clause	Weight
Liability	40
Renewal	20
GoverningLaw	15
Confidentiality carve-outs	15
Termination	10

Formula

Deviation % = (Weighted out-of-bounds ÷ Total checked) × 100

Routing logic:

- <20% → Paralegal (SLA 24h)
- 20–60% → Associate (SLA 8h)
- 60% → Senior (Immediate)

7.) Batch acceptance rules

Objective:

Ensure every batch meets defined quality thresholds before scale-up.

Batch definition:

One document type + one date range.

Required metrics:

Check	Threshold
Naming accuracy	≥ 98%
Metadata completeness	≥ 95%
Critical field error	≤ 1%
Evidence integrity	100%

Sampling plan:

- First 2 batches = 100% review.
- Afterwards = 20% or 10 docs (whichever larger) + all manual reviews.
- Freeze `PlaybookVersion` per batch.

8.) Security & governance controls

Objective:

Embed privilege protection and auditability within the data flow.

Access control:

- Apply *least privilege*.
- Partners/Ops = Owners.
- Working team = Members.
- Visitors = Read-only.
- No external access unless time-boxed via ticket.

AI usage policy:

- Only approved extractor and prompt pack allowed.
- No public LLMs with client data.
- Vendors must not train on firm data.

Change control:

- Every version bump = reason + initials.
- Post to `#ai-changes`.
- Rollback = last valid `/Evidence` version.

Incident handling:

- Errors → `#ai-errors`.
- Triage ≤ 24h.
- Pause workflow until 10–20% spot-check passes.

Audit routine:

- Enable library audit logs.
- Weekly check:
    - External shares = 0
    - Permission drift = 0
    - Sample accuracy ≥90%
    - Retention policy enforced

9.) KPIs and maturity ladder

Core KPIs:

1. Time-to-Triage
2. First-Run Pass Rate
3. Routing Accuracy
4. Playbook Coverage

Health checks:

- Evidence completeness = 100%
- Manual Review rate ≤20%

Day	Goal
14	Repository live + NDA pilot complete
30	Extraction accuracy ≥95%
60	Routing accuracy ≥80%
90	Playbook coverage ≥90% + Governance logs stable

Red flags:

1. Routing drop >10 pts week-on-week
2. Manual Review >20% for 2+ weeks
3. Evidence <100%

Pivot liability caps:

- ≤12 months = Preferred
- ≤18 months = Acceptable
- 18 months = Escalate
- Uncapped = Red flag

10.) Pilot acceptance and scale

Minimum criteria to scale:

1. All batch gates passed
2. Evidence 100%
3. PlaybookVersion frozen
4. Governance logs active

Next step:

- Scale from NDA → MSA or SOW using same folder and column schema.
- Each new doc type = new batch + new acceptance gates.

Book a 15min AI & Data Readiness Check

11.) DIY vs guided implementation

Minimum criteria to scale:

DIY path:

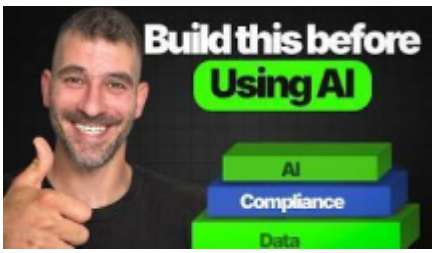
Download the template pack:

- Folder structure (.zip)
- Column schema (.xlsx)
- Extraction prompt JSON
- Triage calculator (.xlsx)

Guided path:

Book a Legal Data Foundation Review with MPL Legal Tech Advisors.

Book a 15min AI & Data Readiness Check



Watch the full breakdown