

AUDIRA: FILE & DATA UPLOAD SCHEMA

Version: 1.0

Purpose: Define the structure, accepted types, metadata requirements, indexing rules, and retrieval process for uploading business documents, APIs, or data sources into Audira's brand brain via RAG or lookup.

Supported File Types

File Type Use Case

| | |
|-------|--|
| .pdf | Policies, tone guides, FAQs, manuals |
| .docx | Training materials, product writeups |
| .txt | Plain scripts, fallback content |
| .csv | Knowledge tables, mapping keys, lookup sets |
| .json | API schemas, endpoint descriptions, config files |
| .xlsx | Process maps, KPIs, CRM mappings |

Images, audio, and video support may be added in later versions (v2.0+).

Upload Metadata Required

Every uploaded file must include:

```
{  
  "business_id": "nova_realty_342",  
  "uploaded_by": "client | consultant | admin",  
  "document_type": "faq | tone_guide | api_doc | other",  
  "visibility": "agent_only | admin_review | full_team",  
  "rag_enabled": true,  
  "indexed": true  
}
```

Indexing Pipeline (for RAG)

1. Pre-processor:

- Converts PDF, DOCX to clean text
- Removes headers/footers and formatting artifacts

2. Chunker:

- Splits content into 200–500 token segments
- Adds document metadata to each chunk

3. Embedder:

- Converts chunks into vector embeddings
- Stores in agent_vectorstore_{ID}

4. Tag Linker:

- Flags related discovery tags (e.g. #fallback_rules, #booking_logic)

5. Access Layer:

- Only agents with permission can query these embeddings
- Auto-updated on file change

Retrieval Rules

- RAG only activates if rag_enabled=true
- Queries are filtered through agent tag logic
- Answers are traceable to source chunk
- Agents must quote or link back to doc if answer is derived from RAG

File Retention & Lifecycle

| Stage | Action |
|--|---|
| Upload | Instant virus scan, metadata registration |
| Index | Added to vector DB and tag graph |
| Update | Re-index triggered, old version archived |
| Delete | Allowed by admin only, audit logged |
| Retention Default = 1 year unless overridden | |

Roles & Access

| Role | Upload | View | Tag | Delete |
|------------|--------|------|-----|--------|
| Client | Yes | Yes | No | No |
| Consultant | Yes | Yes | Yes | No |
| Admin | Yes | Yes | Yes | Yes |

Compliance

- Files containing PII trigger redaction alert
 - Files with compliance relevance auto-link to #compliance_constraints
 - Hashes stored to verify immutability of embedded source
-

Next Document: Agent Simulation Test Kit