# LLM Integration Roadmap for SharePoint-Driven Data Projects

## Overview

### SharePoint Folder Purpose Breakdown

This roadmap outlines the detailed steps needed to build a functional Large Language Model (LLM)-powered assistant that will evolve with the consulting project. Initially, the assistant will provide searchable access to internal documentation (Word docs, PDFs, meeting notes, etc.). In later phases, it will support direct querying of structured, engineered data.

The roadmap includes:

* Folder structure for SharePoint
* File naming conventions and metadata guidance
* Tool selection and integration architecture
* Step-by-step build and deployment stages

## 💪 Project Phases

### **Phase 1: Foundation – SharePoint Structure and Document Indexing**

#### 🌟 Goals

* Create a well-structured SharePoint environment
* Standardize file naming and metadata
* Prepare for LLM ingestion of internal documentation

#### 📂 SharePoint Folder Structure (LLM-Optimized)

/Project\_Root/  
├── 00\_Admin/  
├── 01\_Project\_Plan/  
├── 02\_Client\_Docs/  
│ ├── SOW/  
│ │ └── \_metadata/  
│ └── Internal\_Presentations/  
│ └── \_metadata/  
├── 03\_Internal\_Knowledge/  
│ ├── Meeting\_Notes/  
│ │ └── \_metadata/  
│ ├── Strategy\_Discussions/  
│ │ └── \_metadata/  
│ └── Reference\_PDFs/  
│ └── \_metadata/  
├── 04\_Data/  
│ ├── Raw\_Data\_Dumps/  
│ ├── Cleaned\_Engineering\_Files/  
│ └── Final\_Analysis\_Ready/  
├── 05\_Deliverables/  
│ ├── EDA\_Reports/  
│ └── Final\_Decks/  
└── 99\_Archive/

Folder Purpose

| **00\_Admin/** | Contracts, NDAs, internal instructions, metadata standards, access logs. Used minimally by the LLM, mostly for reference. |
| --- | --- |
| **01\_Project\_Plan/** | Project timelines, Gantt charts, resource plans, and RACI documents. Helps the LLM answer planning and milestone-related queries. |
| **02\_Client\_Docs/** | Files received from or shared with the client. Includes contracts, SOWs, and presentation decks. Enables the LLM to retrieve client inputs, requests, or shared deliverables. |
| ├── **SOW/** | Statements of Work and scope documentation from the client. |
| └── **Internal\_Presentations/** | Presentations created for the client but used internally. |
| **03\_Internal\_Knowledge/** | Internal discussions, decisions, notes, and reference materials. Most valuable for strategic LLM queries. |
| ├── **Meeting\_Notes/** | Notes from both internal and client meetings. |
| ├── **Strategy\_Discussions/** | Decisions around tools, methods, and solution design. |
| └── **Reference\_PDFs/** | Industry research, academic references, vendor documents. |
| **04\_Data/** | All raw, cleaned, and final datasets. Key for structured LLM queries in later phases. |
| ├── **Raw\_Data\_Dumps/** | Original, unprocessed data from ERPs or client exports. |
| ├── **Cleaned\_Engineering\_Files/** | Processed and standardized files for analysis. |
| └── **Final\_Analysis\_Ready/** | Final engineered datasets used for reporting or LLM interaction. |
| **05\_Deliverables/** | All formal outputs and reports delivered to stakeholders. |
| ├── **EDA\_Reports/** | Visuals and write-ups summarizing exploratory findings. |
| └── **Final\_Decks/** | Executive-level summary presentations. |
| **99\_Archive/** | Outdated or versioned-off files. Can be excluded from LLM embeddings unless needed. |

🔸**Note**: Each folder containing document files will include a subfolder named \_metadata/ to store JSON sidecar files. This keeps the working directory clean and separates metadata for easy indexing.

#### 📄 File Naming Convention

YYYY-MM-DD\_[Topic/Meeting/Function]\_[Team/Client]\_vX.ext  
Examples:  
2024-06-10\_Client\_Kickoff\_Notes\_ProjectTeam\_v1.docx  
2024-06-14\_Forecast\_Sync\_Meeting\_Notes\_OpsTeam\_v2.pdf

#### 🏢 Metadata Tagging (via JSON sidecar files)

Each .docx, .pdf, or .xlsx file will include a structured block of text at the top of the document, which can be parsed by automation to generate the JSON sidecar file.

Example block to include at the top of Word, Excel (first tab, top rows), etc.:

Title: Forecast Accuracy Review  
Author: Dan  
Date: 2024-06-14  
Category: Meeting Notes  
Tags: forecasting, demand-planning, accuracy

Python or Power Automate scripts will read this block and generate:

{  
 "title": "Forecast Accuracy Review",  
 "author": "Dan",  
 "date": "2024-06-14",  
 "category": "Meeting Notes",  
 "tags": ["forecasting", "demand-planning", "accuracy"]  
}

Saved in the \_metadata/ subfolder with matching filename:

/Meeting\_Notes/2024-06-14\_Forecast\_Sync\_Meeting\_Notes\_OpsTeam\_v2.pdf  
/Meeting\_Notes/\_metadata/2024-06-14\_Forecast\_Sync\_Meeting\_Notes\_OpsTeam\_v2.json

#### 🛠️ Tools Needed

* SharePoint (with versioning enabled)
* Document templates with metadata block in header
* Script to extract top-of-document metadata and write JSON sidecar files

#### ✅ Deliverables

* SharePoint environment set up with folders and \_metadata/ subfolders
* Word/Excel document templates with built-in metadata headers
* Automation code or process guide to generate JSON sidecars
* Team training on where to input metadata during document creation

## 🧠 Enabling Client Document Indexing

To include documents received from the client:

* Save all files into /02\_Client\_Docs/ with standardized naming conventions
* Request that key context (title, author, purpose, client contact) be added to the document header manually or as part of a submission form
* If no metadata is provided, a team member should fill in the details before LLM ingestion
* JSON sidecar files should be created for each document and stored in the corresponding \_metadata/ folder

For additional robustness, set up a staging folder /02\_Client\_Docs/\_ToBeTagged/ for documents awaiting metadata processing

## 📊 Metadata Categories & Tags (with Definitions)

### 📁 Categories (Use one per document)

| Category | Description |
| --- | --- |
| Meeting Notes | Internal or client-facing meeting documentation |
| Strategy Memo | Internal recommendations, priorities, or architecture decisions |
| Forecast Analysis | Analyses comparing forecast vs actuals, forecast accuracy |
| Inventory Report | Reports detailing inventory position or performance |
| Demand Plan | Final or draft demand planning documentation |
| Supplier Review | Performance evaluations, scorecards, risk assessments |
| Internal Presentation | Slide decks used for internal alignment or team sharing |
| Executive Summary | High-level insight summaries for C-level stakeholders |
| Modeling Document | Descriptions of statistical or ML model structure/logic |
| Data Map | Schema or data lineage documentation |
| EDA Summary | Exploratory data analysis write-ups or charts |
| Client Communication | Files exchanged or produced directly for the client |
| Process Documentation | SOPs, flowcharts, checklists, task ownership docs |

### 🏷️ Tags (Use multiple as needed)

| Tag | Description |
| --- | --- |
| demand-planning | Related to forecasting demand or sales volume |
| supply-planning | Covers supply-side calculations, capacity, or PO planning |
| forecasting | General tag for any forward-looking modeling |
| accuracy | Related to performance vs forecast or target |
| inventory | Any reference to inventory level, movement, or type |
| backorders | Documents that mention or track backorders |
| lead-times | Covers lead time calculations, vendor reliability, etc. |
| suppliers | Specific vendor-related information or evaluation |
| variance-analysis | Statistical comparisons, difference vs baseline or expected |
| data-cleaning | Refers to transformation, filtering, or prep of raw data |
| schema-mapping | Connecting fields across ERPs, schema documentation |
| client-feedback | Notes, reactions, or insight provided by the client |
| kpi-review | Analysis or reporting around KPIs |
| product-alignment | SKU normalization, product hierarchy, naming rules |
| exception-handling | Business rules or workflows for flagged issues |
| python | Scripts or technical logic in Python language |
| dashboard | Deliverables or code related to visualization or UI |
| EDA | Exploratory data analysis tag |
| planning-cycle | Related to monthly/quarterly planning events |
| production | Information about manufacturing flow, WIP, etc. |
| logistics | Transportation, fulfillment, or warehouse handling |
| procurement | Supplier engagement, PO creation, sourcing |

✅ Teams can extend this list as needed. Each new folder or document type should include metadata and an updated metadata\_guide.xlsx in /00\_Admin/Metadata\_Standards/

## 📎 Next Actions

1. Generate a **reference Excel file** with categories/tags + dropdowns.
2. Create a **template ``**\*\* metadata file\*\* you can distribute to the team.
3. Draft a **one-pager metadata guide** to store in /00\_Admin/Metadata\_Standards/

## 🔁 Future Expansion Ideas

* Automate SharePoint sync with Azure Functions
* Index Power BI reports and embedded Excel content
* Create audit logs of LLM queries and outputs for compliance