HOW TO BUILD AN AI-DRIVEN KNOWLEDGE CONTINUITY SYSTEM

FACILITY THOUGHTS

A Practical Internal Roadmap

PREPARED BY

PABLO VELAZQUEZ

How to guide

Who This Document Is For

This guide is intended for facilities and operations leadership, internal IT, data, or automation teams, organizations building continuity capability internally, and global enterprises with varying regulatory and infrastructure constraints. This document assumes no specific software vendor, no requirement to purchase an enterprise AI platform up front, and that incremental implementation is acceptable.

What This System Must Do (Non-Negotiables)

Before discussing tools, the organization must agree on a set of non-negotiable functional requirements. The system must capture operational context from communication, exclude personal, HR, and legal content by design, and produce structured, searchable outputs. It should also support turnover, onboarding, and role consolidation, operate passively without changing how employees work, and be auditable, secure, and defensible. If any of these are removed, the system becomes either unsafe or ineffective.

STEP 1

Step 1: Decide What You Are Capturing (Scope Definition)

It is recommended to start with a narrow scope and not attempt an enterprise-wide ingestion. A good starting point would be to include facilities leadership, regional managers, capital project managers, and maintenance coordinators. It is also crucial to define what constitutes “Operational Content.” Examples include vendor discussions, project updates, quotes and approvals, warranty and compliance communication, and scheduling and escalation emails. Explicit exclusions must be documented, such as HR, legal, medical, performance management, and union correspondence where applicable. This scope definition is a legal and technical input, not a suggestion.

STEP 2

This is the most critical architectural decision.

Step 2: Choose an Email Ingestion Method

Option A: Enterprise Email API (Best Practice, Higher IT Effort)

A server-to-server integration using the organization’s email platform API.

How it works

- IT registers a service application
- OAuth 2.0 client-credentials authentication
- Read-only access
- Permissions restricted to:
    - Specific users OR
    - Specific folders OR
    - Shared operational mailboxes

Why this is preferred

- No passwords stored
- No mailbox delegation
- Fully auditable
- Honors legal holds and retention policies
- Scales cleanly

When to use

- Mature IT organization
- Centralized identity management
- Strong security governance

Cost profile

- Medium engineering effort
- Low long-term operating cost

STEP 2

This is the most critical architectural decision.

Step 2: Choose an Email Ingestion Method

Option B: Transport Rules / Journaling (Low Engineering, High Governance)

Email copies are automatically routed to a controlled internal mailbox based on rules.

How it works

- Transport rules filter by:
  - Sender domain
  - Recipient
  - Keywords
- Copies sent to a continuity mailbox

Pros

- No API permissions required
- Simple to deploy
- Works in restrictive IT environments

Cons

- Can over-capture if rules are loose
- Requires strong filtering discipline

When to use

- Highly regulated environments
- Conservative IT departments
- Early pilots

Cost profile

- Low engineering effort
- Medium governance effort

STEP 2

This is the most critical architectural decision.

Step 2: Choose an Email Ingestion Method

Option C: Company-Owned Processing Mailbox (Budget-Friendly, Flexible)

A centrally managed mailbox created specifically for continuity ingestion.

How it works

- Operational emails are:
  - Auto-forwarded via policy
  - CC'd per process
- Workflow automation pulls from this mailbox

Important This mailbox must be:

- Corporate-owned
- Audited
- Subject to retention rules
- Not a personal account

Why this exists? Some organizations:

- Cannot grant API access
- Operate across multiple email platforms
- Need a clean ingestion boundary

Cost profile

- Low cost
- Moderate operational discipline required

STEP 2

This is the most critical architectural decision.

Step 2: Choose an Email Ingestion Method

Option D: Batch Export (Lowest Cost, Lowest Fidelity)

Periodic exports of email folders or project communications.

When to use

- Very restricted environments
- Offline or air-gapped operations
- Early proof-of-concept work

Tradeoff

Not real-time, but still valuable for continuity.

STEP 3

Step 3: Build the Orchestration Layer

This layer controls how data moves.

Core Responsibilities

- Pull new messages
- Track what has been processed
- Handle failures
- Maintain audit logs

Minimal Viable Setup

- Scheduled job (hourly or daily)
- Email metadata stored in a processing table
- Retry logic for failures

This can be built with:

- Low-code automation tools
- Python scripts
- Internal workflow engines

This layer does not require AI.

STEP 4

Step 4: Sanitize Before AI Touches Anything

This is mandatory.

What Must Be Removed

- Signatures
- Reply chains
- Auto-generated text
- Tracking pixels
- Non-business attachments

What Must Be Redacted

- Personal phone numbers
- Home addresses
- National ID numbers
- Medical references

Why This Matters

- Reduces AI cost
- Improves accuracy
- Protects privacy
- Builds trust with Legal and HR

Sanitization happens before AI processing.

STEP 5

Step 5: Design the AI Processing Pipeline

Do NOT Use One Model for Everything

A robust system separates tasks:

1. Classification
   - What kind of message is this?
2. Entity Extraction
   - Vendors, sites, assets, dates, costs
3. Task Identification
   - Commitments, deadlines, next steps
4. Thread Linking
   - Connect related messages
5. Risk Detection
   - Unresolved issues, delays, repetition

Each task outputs structured JSON, not prose.

STEP 6

Step 6: Define the Data Model Before Scaling

Minimum Required Tables

- Communications
- Vendors
- Sites
- Assets
- Projects
- Tasks
- Relationships

If you skip this step:

- AI output becomes unusable
- Searches become unreliable
- Continuity packets degrade

Semantic Layer

Use embeddings to:

- Link related conversations
- Support natural-language search
- Rebuild narratives across time

STEP 7

Step 7: Generate Continuity Outputs

The Continuity Packet

Automatically generated when:

- An employee exits
- A role changes
- Responsibilities consolidate

Contents:

- Active projects
- Vendor context
- Open risks
- Pending actions
- Historical reasoning

This is the primary business value.

STEP 8

Step 8: Adapt Based on Budget and Maturity

Do NOT Use One Model for Everything

Low Budget / Small Team

- Option C ingestion
- Batch processing
- Local AI or limited API usage
- Manual review

Mid-Tier Organization

- API ingestion
- Automated workflows
- Cloud AI services
- Structured dashboards

Enterprise Scale

- Full API integration
- Role-based dashboards
- Integration with CMMS and ERP
- Automated HR triggers

The architecture is the same. The scale changes.

STEP 9

Step 9: Global and Regional Considerations

Data Residency

- Processing location must match regional laws
- Avoid cross-border transfer unless approved

Labor Laws

- Some regions require notification of automated processing
- Governance must reflect local requirements

Language

- AI models must support multilingual communication
- Entity extraction should be language-agnostic

STEP 10

Step 10: Start with a Paper Pilot

Do NOT Use One Model for Everything

Before production:

- 50–100 historical emails
- Anonymized
- No live connections
- Validate:
  - Accuracy
  - Noise
  - Operational usefulness

This reduces risk and builds confidence.

Final Guidance

This system is not a software purchase.
It is a capability.

Organizations that approach it incrementally, with discipline and governance, will preserve institutional knowledge at scale. Those that skip structure or governance will create risk instead of value.

This roadmap exists to prevent that outcome.