



ORCA paid audit example: RAG documents

Client: Biuro Operacyjne Beta Sp. z o.o. [synthetic]
Process: internal procedure and PDF knowledge assistant
Date: 2026-06-05
Prepared by: Dawid / ORCA

Synthetic example: names, numbers and workflow details are illustrative only. No client data is used.

1. Executive decision

Decision

Prepare data first, then build prototype

Best first automation

Citation-required procedure Q&A assistant

Primary risk

Wrong/outdated source or unauthorized retrieval

Field	Value
Main reason	The business need is strong, but the document corpus has version conflicts and weak access boundaries.
Human approval required?	Yes, before formal/client-facing answers and for policy interpretations.
Suggested next step	7-day document readiness cleanup, then 10-day RAG prototype on a controlled corpus.

2. Current process map

Step	Owner	Input	System/source	Manual work	Output	Pain
1	Operations specialist	Internal question	Teams / email	Searches folders and old PDFs	Informal answer	Slow and inconsistent
2	Process owner	Procedure PDF	SharePoint folders	Checks version and exceptions	Correct procedure fragment	Version confusion
3	Team lead	Draft answer	Email / Teams	Approves interpretation	Final answer	Bottleneck

3. Process fit score

Criterion	Score	Notes
Repeated input	3	80-120 recurring questions per month.
Clear expected output	2	Answer + citation + risk note; some questions need owner approval.
Rules can be written down	2	Procedure hierarchy exists but version policy is unclear.
Human approval is acceptable	3	Formal answers already require approval.
Data access is feasible	2	Documents exist, but roles and folder boundaries need cleanup.
Business pain is visible	3	Search time, repeated questions and inconsistent answers are visible.
Owner is available	2	Process owner can test, but not every day.

Total: 17/21 - possible/strong fit after document readiness cleanup.

4. Corpus and citation readiness

Area	Current state	Risk	Required action
Document versions	Multiple PDFs with unclear current/archived status	High	Mark canonical version and archive old copies.
Metadata	File names contain dates inconsistently	Medium	Add title, owner, version, effective date and role.
Citation quality	Source pages exist but exact clauses are hard to locate	Medium	Store page/section/snippet references.
Access roles	Folders partially map to teams, not decision rights	High	Define role-to-collection map before prototype.

5. Evaluation metrics

Metric	Prototype target	Why it matters
Retrieval recall on gold questions	85%+	Correct source must appear before generation.
Citation support rate	90%+	Claims must be backed by cited snippets.
Abstention on missing evidence	80%+	System should say not enough evidence instead of guessing.
Wrong-version citation rate	<5%	Outdated PDFs must not become authority.
Unauthorized retrieval incidents	0	Role filtering is a production gate.

6. Recommended prototype scope

Goal

Create a controlled assistant that answers operational procedure questions only when it can retrieve supporting source snippets.

Inputs

- 25-40 canonical procedure documents.
- Metadata table: owner, version, effective date, collection, role.
- 30-50 test questions with expected source documents.

Outputs

- Cited answer draft.
- Source snippets and document metadata.
- Abstention reason or exception category.
- Evaluation report across recall, citation support and wrong-version rate.

Out of scope

- Production-wide SharePoint ingestion.
- Autonomous policy interpretation.
- Cross-role retrieval without access control.

7. Next proposal

Recommended paid step	7-day document readiness cleanup + 10-day ORCA Procedure RAG Pilot
Scope	Corpus inventory, canonical-version map, metadata schema, 30-50 question evaluation set, citation-required RAG prototype and evaluation report.
Commercial note	7 500-15 000 PLN netto depending on corpus size, sensitivity and whether first version is local-only or integrated with SharePoint/Teams.