

# PHAROS Training Series - Course 9 "RAG End-to-End: Architecture, Retrieval, Generation and Evaluation"

Contribution ID: 3

Type: **not specified**

## **Hands-on 1: dataset loading, document preparation and first RAG pipeline overview**

*Tuesday, 7 July 2026 11:40 (30 minutes)*

This first hands-on tutorial introduces the Google Colab notebook and the running example used throughout the practical part of the webinar. Participants will load a small Greek-language public-service dataset and inspect how documents are represented before indexing. The session focuses on orientation rather than optimisation: understanding the structure of the corpus, identifying the available fields, examining source information and seeing how raw documents become inputs to a RAG pipeline. Participants will review the full notebook flow, from setup and data loading to chunking, embeddings, retrieval, answer generation and evaluation. They will also inspect intermediate outputs so that the pipeline remains transparent and debuggable. By the end of the session, participants will understand the dataset, the role of metadata, the overall sequence of implementation steps and how the notebook will be used to build a practical RAG system progressively. This foundation helps them follow later implementation choices with confidence during exercises.

**Presenter:** GYFTOPOULOS, Sotiris (ATHENA RC)