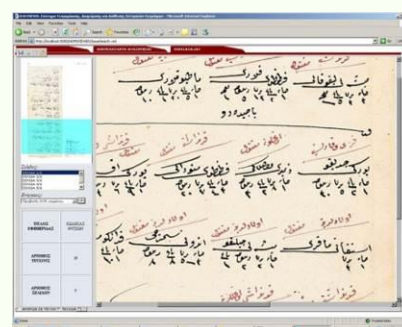# DIATHESIS

## Overview

DIATHESIS is an information system for **documentation**, **management** and **promotion** of **historical documents** that supports both digital library functionality and archival management of the original documents.

The system includes OCR-based **page analysis** and **subject clipping**, subject-level **metadata generation**, semantic **indexing** and multifaceted **classification** of subjects using built-in thesauri aimed at attaining satisfactory levels of efficiency of the classification process, completeness and precision of the retrieved information - supporting both metadata and full-text search - and user-friendly web access.
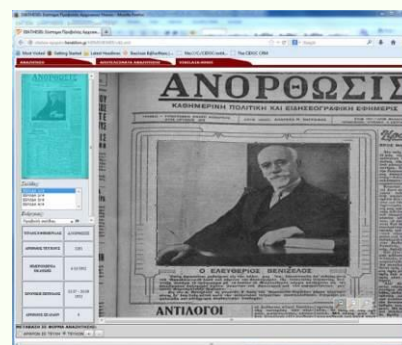
The system implements a "hybrid" form of **classification** and **search** which is based on:

(i) assigning semantic properties, based on the CIDOC-CRM (ISO21127), to specific regions of text that represent a subject of the original document

(ii) full text of the annotated segment of the document produced by an OCR (optical character recognition) session.

The goal of the documentation process is the creation of a **coherent semantic backbone** that can be easily enriched with semantic relations. It is not meant to be a complete semantic structure that includes all the semantic relationships and entities (Actors, Places) described in the text.



*DIATHESIS*



*Vikelaia Municipal Library:*
*Turkish Archive of Heraklion*



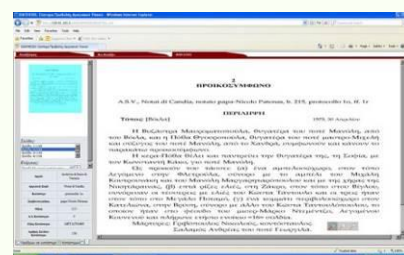*Vikelaia Municipal Library:*
*Newspaper archive*

## Target Domains

The system is suitable for libraries or any other organization that need to archive, document and manage historical documents.
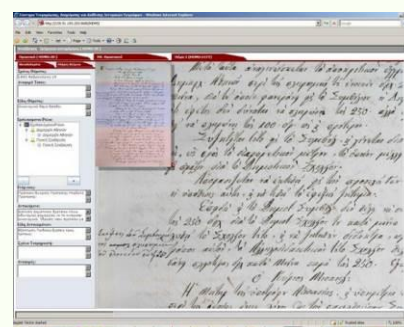
## Description

The DIATHESIS' query interface enables users to conduct searches on a document, as well as on a subject level basis, combining both full text and metadata search capabilities. Queries on the document level are based on conventional metadata assigned automatically to the whole document during the import phase, while queries on the subject level exploit the semantic relationships that have emerged from the documentation phase. The combination of the different query modes provides a semantic filter that greatly improves the precision of the conducted searches. The subject's metadata are based on a robust top level domain ontology (CIDOC-CRM, ISO21127) in order to ensure that the produced knowledge can be inter-exchanged between different institutions. The query result presentation mechanism allows the partial download of the digitized material in order to improve the overall user experience and reduce the download time.

DIATHESIS consists of three lightweight, easily deployable and highly configurable Web applications, namely the administration, the documentation, and the querying applications, which allow data import and monitoring, classification and indexing, and search and presentation respectively.



*ETIA:*
*Archivio di Stato di Venezia,*
*Notai di Candia*



*"Filekpedeftiki Etaireia":*
*Minutes of the Board of Directors and*
*General Assembly*

## Additional Information

The system has been instantiated and tested at:

(i) the Vikelaia Municipal Library of Heraklion, for the archival, documentation and promotion of the archive of Newspapers and Magazines, the Turkish Archive of Heraklion, and the Municipal Archive ("Archio Dimogerontias"). In a first phase 500.000 pages have been digitized and 20% have already been classified and indexed in the system. The system is available online at: *http://vikelaia-epapers.heraklion.gr/BDB/*

(ii) the "Filekpedeftiki Etaireia", an educational non-profit organization founded in 1836 in Greece, for the archival of handwritten manuscripts. The archival material concerns 10 volumes of minutes of the Board of Directors and General Assembly meetings since 1840.

(iii) the newspaper AYGI, for the archival, documentation and promotion of its archives.

(iv) the project "Research and Promotion of Renaissance Cultural Heritage in Crete and Cyprus", *Community Framework Interreg IIIA for Greece and Cyprus*, for historical archives (deeds of contract) available online at: *http://www.medieval-etia-pafos.gr*

**Contact details:** **Maria Theodoridou**
**maria@ics.forth.gr**
*www.ics.forth.gr/isl/cci.html*