[Computational Archival History](#)

Our group develops tools to process and extract text from digitized historical documents in Catalan archives. Using computer vision and natural language processing, we enhance accessibility and searchability in archives, making historical records more usable.

Summary

Vast amounts of historical documents have been preserved in Catalan archives, many of which have been digitized as images. Extracting the text from the images requires a combination of specialist knowledge, cutting-edge computer vision and natural language processing techniques, as well as computer power. The main aim of our group is to create tools for processing and extracting information from these historical documents, to enhance accessibility and searchability in historical archives.

We are working in close collaboration with the Department of Culture of the Generalitat de Catalunya, and with several public and private archives. In one of our projects, we are focusing on the impressive collection of medieval notarial protocols of the Arxiu Històric de Protocols de Barcelona. This collection, unquestionably one of the most noteworthy in Europe, comprises over 3.5K volumes from 1297 to 1500, most of which have been digitized (about 750K images).These documents contain invaluable information about the daily life of the inhabitants of Barcelona, Catalonia and foreigners from Western and Eastern Europe and Africa (merchants and slaves) at the end of the Middle Ages. We are currently developing a system for automatically transcribing this large corpus of medieval documents, rendering their text machine-readable and queryable, making them accessible not only to experts in paleography and medieval history, but also to researchers in other disciplines.

Through our work, we seek to bridge the gap between historical research and modern AI technologies, enabling more efficient exploration, analysis, and preservation of the rich archival heritage of Catalonia.

Objectives

- Develop tools for processing and extracting text from historical documents.
- Enhance accessibility and searchability of Catalan archival records.
- Collaborate with the Department of Culture and various archives.
- Transcribe and digitize medieval notarial protocols.
- Bridge historical research with AI technologies for better analysis and preservation.

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

---

Source URL (retrieved on *2 Abr 2025 - 20:04*): https://www.bsc.es/es/research-development/research-areas/social-simulation/computational-archival-history