

# The e-NDP project: collaborative digital edition of the Chapter registers of Notre-Dame of Paris (1326-1504)

Sergio Torres Aguilar<sup>1</sup>, Julie Claustre<sup>2</sup>, and Darwin Smith<sup>2</sup>

1 Institut für history, University of Luxembourg

2 LaMOP - Université Paris 1 Panthéon-Sorbonne

## 1 Abstract

keywords : *Handwriting Text recognition, Digital Paleography, Chapter registers, Late Medieval manuscripts.*

In this paper we present the e-NDP ([Notre-Dame de Paris et son cloître](#)) project leading with the automatic transcription, text structuration and edition of the digitalized 26 volumes (14 thousand pages) of the capitular registers of Notre-Dame de Paris (1326-1504). The project aims at renewing our knowledge on Notre-Dame de Paris cathedral through the creation of a collaborative digital edition of the registers of its Chapter, the community of canons meeting three times a week on set days to take all administrative, financial and practical decisions pertaining to the cathedral, its estate and the society living in its cloister. The full corpus kept today in the French National Archives, has never been the object of a comprehensive study to understand the workings and history of this urban enclave and community.

The collaborative digital edition of the registers bringing together the Bibliothèque nationale de France, the Archives nationales of France, the École nationale des chartes and the LaMOP-Sorbonne is based on a process of automatic handwriting text recognition (HTR), tested and supervised by scholars, and engineers combining expertise in Medieval history, paleography and artificial intelligence. We apply the new advanced techniques on HTR based on CNN+BiLSTM neural networks with the support of the *Scriptorium* collaborative annotation environment. Our HTR models, published in open access ([here](#)), are currently reaching more of 90% of accuracy on automatic transcription for multilingual manuscripts from the 12th to the 15th centuries. These high quality text transcriptions allow us to perform Natural Language processing (NLP) tasks such as named entity recognition, topic classification and discourse segmentation aiming to transform raw texts into structured texts which are a necessary basis for any digital edition.

In a further step this structured content, already in the form of a database, has been the subject of a first [online publication](#) using the NoSketch engine that allows advanced data mining based on the combination of data, metadata and NLP features. The final step, still in progress, should be the diplomatic editing of the registers on the basis of the current graphic

## 2      **The e-NDP project**

and textual material, but coupled with more refined tools that help in the reading and parsing of the documents such as entity disambiguation, GIS maps and the intertextuality and documentary similarity computing.

This short presentation is also the occasion to show on the one hand the challenges and the working dynamics of a project that brings together the sciences of the erudition and the computational science in a highly specialized work such as the exploitation and transformation of the written heritage by means of computer vision and language processing techniques. On the other hand, it is the occasion to show first-hand research methods focused on a strongly structured edition of a massive corpus, especially to obtain research clues about matters that are arid to the historian, such as inter-institutional relations, political and economic management or about the community life of a central institution of medieval Paris.