

Roman medieval documents meet Machine learning:
a great opportunity for research and teaching

For some years now, our research team (through the 'In codice ratio' and 'Matrices' projects) has been developing tools for the analysis and study of medieval manuscript sources using, in addition to traditional palaeographic and diplomatic skills, machine learning techniques. Our automatic transcription software, applied to the papal registers of Honorius III preserved in the Vatican Apostolic Archive, has given encouraging results; today, with some adaptations, it is applied to the transcription of 12th and 13th century parchments from the fund of Saints Cosmas and Damian preserved in the State Archive of Rome (approximately 200 documents), which have been edited and/or regested only in part. The transcription of the parchments is accompanied by a census of the *scriniarii* involved and a study of their writing practices, which is in turn the basis for a machine learning experiment dedicated to hand recognition and identification (it is known that the same *scriniarii* worked for several clients, and that, during the period under consideration, they themselves affixed the non-autograph signatures of the actors in the document).

This line of research is indeed very promising, given the development of deep learning techniques, such as self-supervised learning, capable of addressing handwriting identification without the need for the census of a huge amount of manuscripts to be used as training samples. Indeed, so far the application of data-driven strategies to palaeography has found a major obstacle in the lack of annotated data, due to the costs, time and expertise required for data labeling. The validity of the experiment we planned for the near future has been confirmed by some preliminary tests conducted on 24 manuscripts from the Vatican Apostolic Library. The meeting in Graz is an opportunity to present our methods, results and challenges, comparing our experience with that of other similar initiatives.

The topics proposed by the CFP that we particularly want to address are:

Machine learning and digital diplomatics
Digital diplomatics in university teaching

Serena Ammirati is Research fellow in Palaeography at Roma Tre University. Her main topics of research are the history of the ancient Latin book (its scripts, formats, layouts and morphologies), the history of written culture in medieval Rome, the Latin book of juridical content from Antiquity through Middle Ages. Since 2016, where she collaborates in projects of Digital Palaeography, in association with the Department of Engineering.

Lorenzo Lastilla is research fellow at the Department of Political Sciences of Roma Tre University, where he works for the ERC project HOLYLAB (<https://holylab-erc.uniroma3.it/>). His main research topic is the application of deep learning solutions on Digital Humanities and Digital Palaeography issues.

Paolo Merialdo is full professor at the Department of Engineering of Roma Tre University. His research interests focus on information extraction, machine learning and explainable AI with applications on Data Management and Digital Humanities issues.

Bibliography

S. Ammirati, *Testi e marginalia in libri prodotti a Roma alla fine dell'alto medioevo e la cultura degli Scriniarii Sanctae Romanae Ecclesiae: alcune riflessioni*, in C. Carbonetti-S. Lucà-M. Signorini (edd.), *Roma e il suo territorio nel medioevo. Le fonti scritte fra tradizione e innovazione. Atti del Convegno internazionale di studio dell'Associazione Italiana Paleografi e Diplomatisti* (Roma, 25-29 settembre 2012), Spoleto 2015, (Studi e ricerche, 6), pp. 343-363

Nieddu, E., Firmani, D., Merialdo, P., & Maiorino, M. (2021). In *Codice Ratio: A crowd-enabled solution for low resource machine transcription of the Vatican Registers*, *Information Processing & Management*, 58, issue 5, september 2021 (doi.org/10.1016/j.ipm.2021.102606)

L. Lastilla, S. Ammirati, D. Firmani, N. Komodakis, P. Merialdo, S. Scardapane, *Self-supervised learning for medieval handwriting identification: a case study from the Vatican Apostolic Library*, *Information processing & management*, 59, issue 3, may 2022 (doi.org/10.1016/j.ipm.2022.102875)