

## **Handling Big Manuscript Data**

Elpida Perdiki and Maria Konstantinidou, Democritus University of Thrace

A recently conceived project on the double tradition of John Chrysostom's works has brought to light questions that a suitably organised VRE could help address. The project in question concerns at least 484 homilies (more than two million words) transmitted in almost 3000 manuscripts. Only a small fraction has been properly edited. Even so, it is impossible to follow the path from the apparatus back to the manuscript's full text. A new transcription of manuscripts is therefore inevitable. Despite the increased accuracy of handwriting OCR tools, they still require to manually transcribe substantial portions of text, a counterproductive process, when dealing with several hundreds of manuscripts. Universal standards in images (as well as in textual data and editorial conventions) need to be established, in order to even begin automatically extracting new data. Further issues involve saving big unstructured data and processing textual data with less than 100% accuracy.

VREs have substantially contributed to better structuring, organising, visualising, linking and sharing data. However, they often arise from needs of a single project and debates on their scope and efficiency is often narrow, fizzling out upon its completion. Could VREs host a community of scholars and institutions (such as libraries) contributing towards a common inquiry? Instead of designing similar digital tools, should we not optimise and expand existing ones to suit a wider range of projects? Manuscripts databases, although increasingly comprehensive in Greek and Latin are not matched by their equivalent in oriental languages (in which valuable translations exist). One hopes to turn to VREs to cross boundaries and bridge the distance.