

The eScriptorium VRE for Manuscript Cultures

P.A. Stokes^{1,2} D. Stökl Ben Ezra^{1,2} B. Kiessling² R. Tissot² H. Gargem²

1. École Pratique des Hautes Études

2. Université Paris Sciences et Lettres

eScriptorium is a VRE and publication framework being developed at Université Paris Sciences et Lettres (PSL) for the study of manuscripts and texts. It comprises the main DH component of Scripta, a project on the history and practices of writing world-wide.

eScriptorium focuses on manual and automatic transcription (using the kraken engine for OCR/HTR), image annotation (using 'deep' structured annotation based on the Archetype framework for palaeography), textual annotation (including post-correction of linguistic markup), and publication. Although still in development, the project has already raised significant questions. The target audience includes researchers in Religious Studies, and indeed the VRE is planned for likely inclusion in a European infrastructure for this domain. However, Religious Studies is an extremely broad field, comprising a very wide range of chronology, geography, and media. eScriptorium focuses on documents and texts, but this already includes writing on parchment, paper, papyrus, stone, clay, palm leaves, bark, and more. Colleagues in Scripta alone work on over fifty different scripts and languages. The disciplinary range includes philology, palaeography, epigraphy, papyrology, and linguistics; extending to Religious Studies adds anthropology, musicology, architecture and much more. But every domain – and every researcher – brings different points of view: research by definition means innovation in methods and even vocabularies, and it often focuses on outliers and exceptions. So how can a VRE be general enough to accommodate these views while having manageable models and interfaces? How do we balance the desire for standardisation and interchange against the complexity and originality of each piece of research? Is it feasible for one VRE to accommodate dozens of rare writing-systems, each with their own conventions and even directionality and not necessarily even represented in Unicode? As well as presenting eScriptorium, this paper will reflect on these challenges and how we might begin to address them.