# "Text-and-Image scholarship for the Renaissance"

**Marie-Luce DEMONET** ***(Centre d'Études Supérieures de la Renaissance, Université François-Rabelais, Tours-CNRS, UMR 6576)***

On July 24, 2009, the first release of the Epistemon text database was included in the BVH (Bibliothèques Virtuelles Humanistes – Humanistic Virtual Libraries, (http://www.bvh.univ-tours.fr), using the XTF search engine : this means that this small French Renaissance corpus (25 texts in HTML and/or XML/TEI format) is incorporated in a website that mainly displays facsimiles of Sixteenth Century books (385 in total) and offers an access to graphics and images extracted from the books. Images containing scenes and motifs are also indexed through the Iconclass thesaurus. Others, such as ornamental letters, are searchable through specific descriptors (the letter itself and its style, the background, the motif). The texts themselves are « cultural heritage » transcriptions: they preserve the original spelling and punctuation, but they are also encoded and processed in order to provide a minimal text analysis with PhiloLogic. For some texts, we try to align the facsimile and its transcription (the work is still in progress), so that the reader can always check the accuracy of the text, and appreciate the page layout. In the near future, this double display will be offered for manuscripts (notarial acts and handwritten material). The main scholarly fields we try to embrace in the same project are vast: book and art history, literature, all realms of Renaissance encyclopedia, and linguistics. This project implies a strong collaboration with computing teams for image analysis: the graphics are extracted with a specific application (AGORA), and classified, thanks to semi-automatic scenarios. Text acquisition should be improved by a heavy scientific investment in character recognition with a new OCR (RETRO), even if, for the first set of texts, we use manual keyboarding. Text mining must overcome the huge obstacle that the variety of spellings offers to the scholar, and the tools of TAL are required to solve the problems of ambiguity (with Dissimilog and Analog software).
Collaborations are worldwide: the basic transcriptions are outsourced in Southern Asia; the software PhiloLogic processes the text documents from Chicago University; the Iconclass thesaurus browses the keywords for the graphics from the Netherlands; students of the Indian Institute of Information and Technology help solving the OCR difficulties and managing the digital library itself. Lately, the BVH project joined the Europeana consortium in the work package dealing with semantic web.

Renaissance scholarship and history of printing benefit a considerable amount from the many capacities we can gather, and also from the models provided by other similar projects using an XML/TEI architecture (Newton website, DTA Archive…). What impact on scholarly knowledge can we hope to have for the present? Students, academics, and all kinds of users find it normal to have access to images and to texts, but the difficulties of actually setting up this double access must not be underestimated. Technologies differ, and there are still serious gaps between image processing, indexing, and text analysis. The book —or more generally, the document— can be considered as an object belonging to cultural heritage, and also a set of signs bearing a large number of significations that are not only textual.

In fact, the new knowledge extends to the research itself within the improvement of digital access: in literature, TEI encoding forces us to revise the decisions about the « best » editions of classics, asking serious questions about the definitions of genres, about real or fictitious names, about the boarders between literary and non-literary texts, and solving conflicts between a physical ontology (the page), and a logical one (the word and the sentence). In history of linguistics, a new knowledge emerges about evolution of semantic distribution, word order, alternate spellings, punctuation and morphosyntax. In book history, the study of the OCR

patterns leads to discoveries concerning typography, counterfeits, sets of ornamental letters, and reuse of graphics. Visualizations offer new hypotheses about the circulation of texts, scholarly networks, and localizations.

Digitization is not only a way of offering facsimiles and new sophisticated editions, of acquiring new technical knowledge: the process itself also generates an actual improvement within the domains of humanities, a kind of knowledge "mining" that goes far beyond our expectations.

Pr. Marie-Luce Demonet
Head of the "Bibliothèques Virtuelles Humanistes"
marie-luce.demonet@univ-tours.fr
http://www.bvh.univ-tours.fr
Vice – director of the Centre d'Etudes Supérieures de la Renaissance, UMR-CNRS
Université François-Rabelais, Tours
59, rue Néricault-Destouches, BP 1328, 37013 Tours Cedex, France
33 (0)2 47 36 77 80 fax 33 (0)2 47 36 77 62