

Collaboration as a Key to The Setting Up of a ‘Digital Registry for Chinese Publications’ - for Digitized Chinese Books, Journals and Newspapers

*Rita WONG, Colin STOREY

University Library System, The Chinese University of Hong Kong, Hong Kong

rita-wong@cuhk.edu.hk

Chinese culture has a long history of written records in paper form. There are at least 3 million Chinese rare books (published from 960 to 1911 A.D.) located in Mainland China and Taiwan, over 300,000 volumes in Europe and 900,000 volumes in the United States. Recently, libraries and commercial entities have initiated a massive digitization of Chinese publications. This calls for a ‘Digital Registry’ to record what has already been digitized, and what is going to be digitized, so that the technology and experience can be shared and the duplication of effort avoided.

The DLF/OCLC Registry of Digital Masters developed by Online Computer Library Center (OCLC) and the Digital Library Federation (DLF) is a large scale digital registry mainly for digitized Western books. Currently, searching in public mode with Chinese characters is not possible. The European Library, a portal for 45 national libraries of Europe, similarly provides retrieval of digital objects, but with limited Chinese publication content. The proposed ‘Digital Registry for Chinese Publications’ will serve a similar function to the DLF/OCLC Registry of Digital Masters and will be mounted in a neutral place, such as Hong Kong. The proposed prototype will embrace Chinese rare books in stitch-bound format and modern text. A comprehensive survey will be conducted to ascertain *which* libraries or other bodies are digitizing and *what* they are digitizing (i.e. books, journals or newspapers.) The metadata standards employed will be compared. Before a metadata structure for the ‘Digital Registry’ is finalized, reference will be made to the metadata standards for Chinese rare books proposed by Mainland China and Taiwan.

One of the functions of the ‘Digital Registry’ will be the listing of internal codes of Chinese characters used by each library or body in digitization if electronic text is included. Different internal codes, such as GB, Big 5, CCCII or EACC, are used in Mainland China, Taiwan, Hong Kong and the Western world. The ‘Digital Registry’ will also record what Chinese characters are not found in the latest version of Unicode Version 4.0 developed by the Unicode Consortium. The recording of such infrequently used Chinese characters will assist libraries and other units in identifying the occurrence and frequency of specific characters with the ultimate aim of inclusion in the Unicode Standard. Other important data related to digitization will be recorded in the ‘Digital Registry’, such as format, resolution and storage device.