

University of Denver

## Digital Commons @ DU

---

Library and Information Science: Faculty  
Publications

Library and Information Science: Faculty  
Scholarship

---

2015

# Multilingual Metadata for Cultural Heritage Materials: The Case of the Tse-Tsung Chow Collection of Chinese Scrolls and Fan Paintings

Krystyna K. Matusiak  
*University of Denver*, [krystyna.matusiak@du.edu](mailto:krystyna.matusiak@du.edu)

Ling Meng  
*University of Wisconsin - Milwaukee*

Ewa Barczyk  
*University of Wisconsin - Milwaukee*, [ewa@uwm.edu](mailto:ewa@uwm.edu)

Chia-Jung Shih  
*University of Wisconsin - Milwaukee*

Follow this and additional works at: [https://digitalcommons.du.edu/lis\\_facpub](https://digitalcommons.du.edu/lis_facpub)



Part of the [Cataloging and Metadata Commons](#)



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

---

### Recommended Citation

Matusiak, K.K., Meng, L., Barczyk, E., & Shih, C.J. (2015). Multilingual metadata for cultural heritage materials: The case of the Tse-Tsung Chow Collection of Chinese scrolls and fan paintings. *The Electronic Library*, 33(1), 136-151. <https://doi.org/10.1108/EL-08-2013-0141>



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

This Article is brought to you for free and open access by the Library and Information Science: Faculty Scholarship at Digital Commons @ DU. It has been accepted for inclusion in Library and Information Science: Faculty Publications by an authorized administrator of Digital Commons @ DU. For more information, please contact [jennifer.cox@du.edu](mailto:jennifer.cox@du.edu), [dig-commons@du.edu](mailto:dig-commons@du.edu).

---

## Multilingual Metadata for Cultural Heritage Materials: The Case of the Tse-Tsung Chow Collection of Chinese Scrolls and Fan Paintings

### Comments

This is a postprint version of the article published in the *Electronic Library*. The final version may be found at <https://doi.org/10.1108/EL-08-2013-0141>.

### Publication Statement

Copyright held by the author or publisher. User is responsible for all copyright compliance.

This is a postprint version of the article published in the *Electronic Library*. The final version may be found at <https://doi.org/10.1108/EL-08-2013-0141>.

Preferred citation:

Matusiak, K.K., Meng, L., Barczyk, E., & Shih, C.J. (2015). Multilingual metadata for cultural heritage materials: The case of the Tse-Tsung Chow Collection of Chinese scrolls and fan paintings. *The Electronic Library*, 33(1), 136 - 151.

### **Multilingual Metadata for Cultural Heritage Materials: The Case of the *Tse-Tsung Chow Collection of Chinese Scrolls and Fan Paintings***

Krystyna K. Matusiak

Library & Information Science Program, Morgridge College of Education,  
University of Denver, Denver, Colorado, USA, and

Ling Meng, Ewa Barczyk and Chia-Jung Shih  
UWM Libraries, University of Wisconsin-Milwaukee, Milwaukee,  
Wisconsin, USA

#### **Abstract**

**Purpose** – The purpose of this paper is to explore multilingual access in digital libraries and to present a case study of creating bilingual metadata records for the Tse-Tsung Chow Collection of Chinese Scrolls and Fan Paintings. The project, undertaken at the University of Wisconsin-Milwaukee Libraries, provides access to digital copies of calligraphic and painted Chinese scrolls and fans from the collection donated by Prof Tse-Tsung Chow (Cezong Zhou).

**Design/methodology/approach** – This paper examines the current approaches to multilingual indexing and retrieval in digital collections and presents a model of creating bilingual parallel records that combines translation with controlled vocabulary mapping.

**Findings** – Creating multilingual metadata records for cultural heritage materials is in an early phase of development. Bilingual metadata created through human translation and controlled vocabulary mapping represents one of the approaches to multilingual access in digital libraries. Multilingual indexing of collections of international origin addresses the linguistic needs of the target audience, connects the digitized objects to their respective cultures and contributes to richer descriptive records. The approach that relies on human translation and research can be undertaken in small-scale digitization projects of rare cultural heritage materials. Language and subject expertise are required to create bilingual metadata records.

**Research limitations/implications** – This paper presents the results of a case study. The approach to multilingual access that involves research, and it relies on human translation that can only be undertaken in small-scale projects.

**Practical implications** – This case study of creating parallel records with a combination of translation and vocabulary mapping can be useful for designing similar bilingual digital collections.

**Social implications** – This paper also discusses the obligations of holding institutions in undertaking digital conversion of the cultural heritage materials that originated in other countries, especially in regard to providing metadata records that reflect the language of the originating community.

**Originality/value** – The research and practice in multilingual indexing of cultural heritage materials are very limited. There are no standardized models of how to approach building multilingual digital

## 1. Introduction

Digitization projects bring to light valuable cultural heritage materials held in special collections, archives, and museums. Digital collections, created as a result of digitization, offer users unprecedented access to rare textual, visual, and multimedia resources from diverse cultures. Special collections of interest to international users often originated in other countries, but happen to be housed in American libraries as a result of the donations from scholars affiliated with U.S. universities and colleges. In many cases, these unique and rare materials represent the cultural heritage of other nations and are relevant to speakers of languages other than English. Digitization offers an opportunity to expand access and to share these collections with users in other countries. The international coverage and the relevance to users outside the United States, however, pose questions about presenting and indexing the non-English content and facilitating access in the multilingual environment

Digital libraries (DLs) have evolved into multimedia and multicultural entities, as Borgman (1997) predicted in the early phase of DL development, but they are rarely multilingual. This is one of the promises of DLs that has not yet been entirely fulfilled despite over a decade of research on multilingual information access (Diekema, 2012). User community on the Web is global and multilingual and, as recent research indicates, users expect multilingual access in digital libraries (Clough & Eleta, 2010; Wu *et al.*, 2012). Herold (2010) in analyzing the patterns of user access in digital collections with content of international origin finds a significantly higher number of searchers from non-United States locations. Not surprisingly, a digital collection documenting the cultures of Malaysia, for example, has a high number of users coming from that region (Herold, 2010). The developers of digital collections with source materials of international origin or with non-English content have begun to consider the needs of global users and to address the complexity of multilingual access in DLs (Gauthereau-Bryson *et al.*, 2011; Herold, 2010; Kupietzky, 2007).

The choice of language for creating metadata records and for presenting content raises a number of questions for librarians building digital collections. How do you describe digitized objects if they represent the heritage of users in other cultures and are of potential scholarly interest to researchers worldwide? What should be the language of metadata records in digital collections with diverse cultural and linguistic content? Do you translate the content of original source materials if the text is in language(s) other than English? The language of description and access points in metadata records does matter as it is a key to resource discovery. Information retrieval of digital objects is very much a linguistic endeavor. Monolingual metadata in English can create barriers to discovery and use for international users, or as Herold (2010) points out, will limit access to elite scholars who are fluent in English. The tools offered by search engines, such as Google Translate or Bing Translator provide the option of translating English text into a user's native language, but the results of automated translation, especially without at least some level of human editing are rather crude, inadequately representing descriptive records and content of literary and artistic works. Despite the evidence that the audience of digital collections is global and multilingual, descriptions of digital objects as well as the tools for their access and retrieval remain largely monolingual. Multilingual access remains a goal for all DLs,

but the issue is particularly pertinent to digital collections presenting source materials that originated in other countries.

This paper describes the process of creating bilingual metadata records for the *Tse-Tsung Chow Collection of Chinese Scrolls and Fan Paintings*. The project, undertaken at the University of Wisconsin-Milwaukee (UWM) Libraries, provides access to digital copies of Chinese calligraphic scrolls and painted fans from the collection donated by Prof. Tse-Tsung Chow (Cezong Zhou). The original collection holds special meaning among speakers of the Chinese language and to those who are interested in Chinese cultural heritage. Thus, it seemed proper to make the digital representation accessible in the language in which the original artifacts were created. The search for models and similar projects demonstrated that the number of multilingual digital collections is relatively small and their approaches for facilitating access in multiple languages vary. Multilingual access is about crossing the language barrier between users and available information resources, but this goal can be attempted in a number of ways in the current digital library environment.

## **2. Background: Multilingual Access in Digital Libraries**

Multilingual access to digital objects in DLs has been an area of growing research in recent years, but thus far it has limited applications in practice (Diekema, 2012). Several models and technological solutions are available to provide multilingual access in DLs. They vary from human translation of metadata records and the development of multilingual controlled vocabularies to automated methods, such as machine translation (MT) and other cross-language information retrieval methods (CLIR). Chen and Bao (2009) provide an overview of cross-language information retrieval and its possible applications for the development of multilingual digital libraries. The authors note that search engines, such as Google or Bing, recognize the linguistic diversity of Web users and offer support for cross language searching. In the DL environment, however, there seems to be a gap between research efforts and the practice community. As one of the studies indicates, automated approaches like MT or CLIR have not been implemented in the existing multilingual DLs beyond the experimental stage (Weaver-Budzise *et al.*, 2012).

The research on multilingual metadata and indexing vocabularies is particularly relevant to this project and other collections of cultural heritage materials with content in more than one language. Multilingual searching of metadata records requires that either the metadata are provided in multiple languages that users are able to search or user query is translated into language of the metadata. Metadata records, especially in the collections of visual and other non-textual resources, provide major points of access, if not the only ones. The need to create multilingual tools for the indexing of visual resources has been recognized in the museum environment where objects can be registered in several languages (Kupietzky, 2007). In addition, unlike print textual materials, visual resources don't have a linguistic source of description, such as a title page or a table of contents. The multilingual context where the language of a user query can be different than the indexing language adds to the complexity of the visual resource discovery. Ménard (2010) points out the double challenge in searching for images in the multilingual context where a user query has to correspond to the textual descriptors in metadata records and the language of the query has to match the language used in image indexing. The researcher finds a lower user satisfaction in image retrieval where a query needs to be translated. Ménard (2012) considers creating multilingual indexing tools, such as a bilingual

taxonomy for English and French speakers in her research project TIIARA (Taxonomy Image Indexing And Retrieval) as a solution for enhancing image retrieval in the multilingual context. A bilingual lexicon with English and Hebrew terms was also built for indexing of cultural artifacts in the Israel Museum in Jerusalem (Kupietzky, 2007). The Getty vocabularies, the thesauri developed by the Getty Research Institute and used widely for indexing of visual resources, are increasingly becoming multilingual. The Art and Architecture Thesaurus (AAT), for example, includes translations of English terms into Spanish, Chinese, Dutch, Italian, and French (Harpring, 2013).

The development of new bilingual taxonomies requires a tremendous effort, cost, and domain knowledge. The other approach aims at establishing interoperability between the existing controlled vocabulary tools in different languages through vocabulary mapping (Chen *et al.*, 2011; Chen & Chen, 2012). The vocabulary mapping methodology was developed as part of an effort to share the rich resources of Chinese cultural heritage in TELDAP (Taiwan e-Learning and Digital Archives Program) with an international audience. The proposed framework includes four modules: translation, mapping, localization, and creation. This approach addresses the issues of cultural differences reflected in the respective languages, English and Chinese. The model developed at TELDAP is quite unique and it presents a tremendous potential for the world of DLs where interoperability is one of the key standards. Multilingual controlled vocabularies or vocabulary mapping address the language of subject descriptors. Other metadata elements that use natural language, such as title or description require human or machine translation.

Machine translation (MT) is a subfield of Artificial Intelligence. It automates the process of language translation by analyzing and understanding information in one language and expressing it in another language. MT offers several solutions for multilingual access in digital libraries with automated translations of user queries, documents, or metadata records (Chen and Bao, 2009; Jones *et al.*, 2007). Translation of metadata records can be conducted by using freely available translation applications on the Web, like Google Translate or Bing Translator (Chen and Bao, 2009; Chen *et al.*, 2012). MT offers the advantage of translating large amounts of text instantly and at a low cost, but the accuracy of translation is limited. Jones *et al.* (2007) also note the challenge of automated translation of metadata records that typically lack the linguistic structure and redundancy of full text documents and have a high density of search terms. The research on the machine translation of metadata records in the context of DLs has focused on improving the quality of translation by combining standard automated translation with domain-specific lexicon (Jones *et al.*, 2007). Automated translation can also be augmented by human evaluation to improve their accuracy and cultural appropriateness (Chen *et al.*, 2011). The research on machine translation of metadata records is in its early phase, but it offers a significant promise for multilingual access to a large body of metadata records in existing DLs.

Human translation represents a traditional approach to providing access to texts and records in the multilingual context. It contrast to automated methods, it renders more accurate metadata and captures the cultural nuances but is resource and cost intensive. In current practice, the actual multilingual digital collections are few and they rely mainly on human translation of metadata records or original documents. The most notable example is the World Digital Library (WDL), a collaborative project undertaken by UNESCO, Library of Congress, and a significant number of libraries worldwide (Van Oudenaren, 2010). The WDL provides multilingual access to unique resources from many different cultures through item-level metadata records translated consistently into seven languages. The collection interface is also localized and available in seven languages. A different model of multilingual access is represented by the Our Americas

Archive Partnership (OOAP), a collaborative project between Rice University, the University of Maryland, and Instituto Mora in Mexico. The collection provides access to the translated content of primary documents in French, Portuguese, and Spanish, but metadata records are monolingual and created in English at this point. The source documents and their “born-digital translations” are searchable and presented within the same collection. Translation of metadata records represents a future goal of making the Our Americas Archive Partnership truly multilingual (Gauthereau-Bryson *et al.*, 2011).

The project of creating bilingual records for the *Tse-Tsung Chow Collection* is small in scale and represents a practical approach where human translation of metadata records was combined with controlled vocabulary mapping.

### **3. The *Tse-Tsung Chow Collection of Chinese Scrolls and Fan Paintings***

Tse-Tsung Chow (1916-2007), also known as Cezong Zhou, was a historian, a poet, and a professor of the Department of East Asian Languages and Literature at University of Wisconsin-Madison. His book, *The May Fourth Movement: Intellectual Revolution in Modern China* (1960) is regarded as being one of the first significant works on modern China. Tse-Tsung Chow also collected calligraphic and painted scrolls and fan paintings in the spirit of the ancient Chinese tradition where art collecting is considered a form of communication. According to Kraus (1991), the exchanging of artwork, such as calligraphic poems, was viewed in the Qing dynasty as a practice of combining power and aesthetics. In 2005, Tse-Tsung Chow (who died in 2007) and his wife Nancy Wu Chow donated over 120 scrolls and fan paintings from his collection to the University of Wisconsin-Milwaukee (UWM) Libraries. To continue the tradition of exchange and sharing, Tse-Tsung Chow asked the UWM Libraries to provide access to the collection online.

The collection offers primary examples of Chinese culture spanning over two hundred years with calligraphic and painted scrolls and fans ranging from the 18th through the 20th centuries (between the Qing dynasty and the beginning of Republic of China). The *Tse-Tsung Chow Collection* provides an insight into the variety of styles and techniques of Chinese scroll and fan painting. The original collection consists of 129 items: 23 couplets (multi-item objects) and 83 single-item objects. Single items include individual hanging scrolls, fan paintings, calligraphic fans, and banners. Figure 1 demonstrates an example of a couplet, while Figure 2 provides an example of a fan painting. The original collection, however, includes only a limited amount of descriptive information. Therefore, extensive research had to be undertaken to index the objects selected for the online collection.



Figure1. Rengong Liang's (Qichao Liang) calligraphic couplet - 梁任公書集聯.

#### 4. Digitizing the Tse-Tsung Chow Collection of Chinese Scrolls and Fan Paintings

The digitization project was undertaken in the spring of 2011 and consisted of two phases. In the first, exploratory phase of the project, the librarians researched the models of multilingual access, created a bilingual metadata template, and selected items for a pilot digital collection. The first phase of the project provided an opportunity to research the original source items and present the first batch of bilingual records online. CONTENTdm, a digital collection management system, was a natural choice as a delivery platform, as it is being used at the UWM Libraries for presenting other digital collections and it supports Unicode, necessary for presenting Chinese characters.

In the first phase of the projects 50 objects (60 items) were digitized and uploaded to the online collection with bilingual records. The remaining items were added in the second phase of the project. The online collection was completed in the fall of 2012. It consists of 98 records, representing 121 items, 23 couplets (multi-item objects) and 75 single-item objects. A small number of scrolls from the collection were not digitized because the original items were too fragile for scanning. The online collection created as a result of the project is open to the public and available online at: <http://www4.uwm.edu/libraries/digilib/scroll/>.





Figure 2. Handi Chen's fan painting with ink-bamboo - 陳漢第墨竹扇面畫.

## 5. Building Bilingual Records

The process of building bilingual records for the collection involved customization of a default Dublin Core metadata template; transcribing the textual content of calligraphic scrolls and translating it into English; creating metadata records and translating the values from Chinese into English; and mapping controlled vocabulary terms. Other options, discussed in the research literature, such as machine metadata translation, were not feasible since the original collection included very little descriptive information, and even when Chinese language text was available the results of automated translation were far from satisfactory. A test with machine translation (MT) was conducted using Google Translate. Selected metadata fields with Chinese language text were translated using automated approach and compared with human translation. Table 1 presents the comparative results and demonstrates the limitations of machine translation in its current stage of development. The project team selected human translation as the best approach for capturing linguistic and cultural characteristics of original materials, and strived for creating parallel records, taking into consideration the differences between the languages, cultures, and literary styles. The bilingual indexing process was accompanied by extensive research necessary to identify artists' names and their backgrounds and to recognize the scripts and seals.

Insert Table 1. A comparison of human and machine translation of the selected metadata fields in a record in the *Tse-Tsung Chow Collection*. The record is available at: <http://collections.lib.uwm.edu/cdm/ref/collection/scroll/id/132>. The machine translation was performed with Google Translate.

Metadata Field	Chinese Language Text	Human Translation into English	Machine Translation into English
Title	汪中書康鼎銘	Zhong Wang's calligraphic hanging scroll with inscription of Kang Ding	Kangding Ming Wang in the book
Artist Bio	汪中，字容甫(容父)，號頌父，江蘇人，為清朝著名文學家及史學家。汪中熟讀經史百家書籍，擅長駢文，其駢文作品「哀鹽船文」最被後人所推崇。汪中研究墨子，善治學，文學創作豐富，被歸類為揚州學派。	Zhong Wang, known as Rongfu (courtesy name), was born in Jiangsu, China. Wang is a renowned litterateur and historian during the era of the Qianlong Emperor (1711-1799). Wang is interested in classics, history and philosophy. He is an expert of Mozi's philosophy (Mozi, ca. 470 BC – ca. 391 BC, a Chinese philosopher at early Warring States Period) and excelled at research on classics. His style of research is categorized as “Yangzhou School” (an academic school originated from Yanwu Gu (1613-1682)). He also specialized in parallel prose. His work, “Ai Yan Chuan Wen” was highly praised by Shijun Huang (1696-1773, a well-known scholar specializing in calligraphy, poetry and history).	Wang, the word Fu Yung (Yung father), No. Chung Father, Jiangsu, the Qing Dynasty famous writer and historian. Wang in the familiar classics, history, one hundred books, good prose, its prose work "sorrow salt Boat Man" the most respected posterity. Wang studied Mo, good scholarship, literature-rich, is classified as Yangzhou School.
Main Text	唯三月初吉甲戌，王康宮，榮白內右康，王命死辭王家，右幽黃，康拜稽首，敢對揚丙子不顯休，用乍朕文考釐白寶尊鼎，	During an era of Western Zhou (1046-771BCE), the emperor was in his imperial palace on an auspicious day of 934BCE. Duke, Rong, lead a person, Kang, to his space. The emperor issued an order and said, “I	Ji Jia Xu Wei early March, Wang Kang Palace, within the right-wing white Kang, Wang Ming Wang died speech, the right secluded yellow, Awn Jishou, dare not significant Hugh Young Rat, at first I used the

### 5.1. Customization of Metadata Template

The default Dublin Core metadata template, available in CONTENTdm, was customized to address the unique features of the collection and to build bilingual fields (see Figure 3 for a sample). Given the artistic expressions and the language of the collection, the project staff conducted some research in calligraphy and Chinese paintings. The collections of the National Palace Museum in Taiwan were studied in the process. New fields, such as Collector, Period, Seal Content were introduced to the template.

	Field name	DC map	Data type	Large	Search	Hide	Required	Vocab
1	Title/標題	Title	Text	No	Yes	No	Yes	No
2	Artist	Creator	Text	No	Yes	No	No	Yes
3	作者	Creator	Text	No	Yes	No	No	Yes
4	Artist Bio	Description	Text	Yes	No	No	No	No
5	作者生平	Description	Text	Yes	No	No	No	No
6	Date of Creation	Date-Created	Text	No	Yes	No	No	No
7	創作日期	Date-Created	Text	No	Yes	No	No	No
8	Period	Date	Text	No	Yes	No	No	No
9	年代	Date	Text	No	Yes	No	No	No
10	Collector	Contributors	Text	No	Yes	No	No	Yes
11	收藏者	Contributors	Text	No	Yes	No	No	Yes
12	Description	Description	Text	Yes	No	No	No	No
13	描述	Description	Text	Yes	No	No	No	No
14	Main Text	Description	Text	Yes	Yes	No	No	No
15	釋文	Description	Text	Yes	Yes	No	No	No
16	Other Text	Description	Text	Yes	Yes	No	No	No
17	題跋與款識	Description	Text	Yes	Yes	No	No	No
18	Seal Content	Description	Text	No	Yes	No	No	No
19	印記	Description	Text	No	Yes	No	No	No
20	Language	Language	Text	No	Yes	No	No	No
21	語言	Language	Text	No	Yes	No	No	No
22	Subject AAT	Subject	Text	No	Yes	No	No	Yes
23	主題 AAT-Taiwan	Subject	Text	No	Yes	No	No	Yes

Insert Figure 3. Selected bilingual fields in the customized CONTENTdm metadata template. The entire template includes 51 fields.

The process of translating field names in the metadata template was bi-directional. The standard Dublin Core fields, such as Title or Date, were translated from English to Chinese; while the unique ones were first created in Chinese and then translated into English. The final template includes 51 fields with 25 fields for each language and one shared bilingual Title field. The project team decided to use one field for Title to have the English and Chinese version displayed in the Results page along with thumbnail images. The names of fields were customized and mapped to Dublin Core to ensure interoperability and to enable cross-collection searching and metadata harvesting.

## 5.2 Translation

The original language of the collection is Chinese. The calligraphic paintings, included in the collection, were created in a variety of Chinese language scripts. Thus, Chinese was chosen as a source language for metadata records, which were then translated into English. The project team included two native speakers of Chinese with background in Chinese literature and art who conducted most of the research, created metadata records in Chinese, and translated them into English.

The bilingual data accompany the images of scrolls and paintings in the digital collection, with Chinese text and English translation interfiled in the same metadata record (see Figure 4 for an example of a bilingual record). Translation proved to be the most complex and challenging aspect of the project due to a variety of scripts and literary styles and forms represented in the collection, but also because of the inherent difficulty in translating the Chinese written language into English. Three metadata fields, author's name, the main literary text, and seal inscriptions proved to be particularly challenging.

<b>Main Text</b>	Spring orchids and autumn chrysanthus blossom every year and remain fresh all the time. Bright moonlight and crystal frosted dew imply changes of four seasons and indicate the passing of time.
釋文	春菊秋蘭, 英華靡絕。明月白露, 光陰往來。
<b>Other Text</b>	Preface: dedicated to Mei Cai; Signature: Liang Qichao in the winter of 1927.
題跋與款識	款識: 美才仁兄清屬, 丁卯季冬, 梁啟超
<b>Seal Content</b>	1) zhi yu zhi shang; 2) ma ti qiu shui hu wei chun bing; 3) Chow Tse Tsung; 4) Liang Xinhui; 5) Qichao private seal
印記	1) 止於至賞; 2) 馬蹄秋水虎尾春冰; 3) 周策縱; 4) 新會梁氏; 5) 啟超私印
<b>Language</b>	ch
語言	中文
<b>Subject AAT</b>	Calligraphy (visual work) Hanging scrolls Poetry
主題 AAT-Taiwan	書法(視覺作品) 立軸 詩
<b>Subject LC</b>	Liang, Qichao, 1873-1929
<b>Coverage</b>	China
創作地	中國

Insert Figure 4. Selected bilingual fields in the metadata record of the calligraphic couplet, presented in Figure 1. The full metadata record is available at: <http://collections.lib.uwm.edu/cdm/ref/collection/scroll/id/36>

The project team adopted two approaches to representing the text and Chinese characters in the English translation. The literary text depicted in the scrolls was paraphrased and its meaning was translated into English. However, proper names and the terms that appeared in the seals but had no English equivalents required a different solution. Following a standard practice in representing the Chinese characters in Latin alphabets, the terms were transcribed from pronunciation. The difficulty in transcribing historical texts lies in the fact that the pronunciation of Chinese characters has changed over the centuries. Standard Chinese (or Modern Standard Chinese) was chosen to encode the data consistently. Since different phonetic transcription systems exist to represent Chinese characters the most popular system, Hanyu Pinyin, which fulfills the ISO 7098 standard, was selected.

The seal inscriptions that indicate the changing ownership of the scrolls and intellectual contributions required special attention in the translation process. The original text had to be transcribed and in some cases deciphered before it could be translated. An independent Chinese researcher, Mr. Xu, volunteered to assist the team in the task of recognizing some scripts and seals.

The translation of the literary texts represented a different set of challenges due to a variety of scripts and literary styles. The scrolls include text in seal script, bronze script, oracle bone script, semi-cursive script, standard script, running cursive script, and cursive script. Figure 5 provides an example of cursive script, a style of Chinese calligraphy which is written quickly and can be quite difficult to read. The calligraphic scrolls represent a wide range of literary forms from poems, odes, eulogies, and mottos to short essays and posed an additional challenge in the translation process. It was very difficult to find corresponding literary styles between ancient Chinese and modern English. It was also impossible to keep the rhymes, but the translators tried to recreate the basic structure of the forms and to maintain the original paragraph patterns. The goal was to convey the meaning and give users a sense of the original item's artistic form.



Insert Figure5. Scroll prints with Xu Zhang's cursive calligraphy - 張旭草書拓本

Most of the original literary pieces were written in Classical Chinese (Literary Chinese), a traditional style of written Chinese based on the grammar and vocabularies of ancient Chinese, which is different from any modern spoken form of Chinese. It is very difficult to translate directly from Classical Chinese to English. The team's solution was to translate from Classical Chinese to modern Chinese first, and then from modern Chinese to English.

### 5.3. Controlled Vocabulary Tools and Mapping of Chinese and English Terms

Multiple controlled vocabulary tools as well as natural language were used to record metadata values. Most of the controlled vocabulary tools proved to be in English. The project team identified only one bilingual thesaurus - AAT-Taiwan (Art & Architecture Thesaurus – Taiwan). Natural language was used for descriptive fields, such as Main Text, Artist Biography, and Seal Content. The main text of calligraphic scrolls was transcribed and translated into English using natural language. Other fields, such as Titles, Artist Bio, and Seal Content were written in Chinese and then translated into English.

The project team made an effort to use controlled vocabulary for subject headings and to map the terms rather than translate them. The attempt to create parallel records in Chinese and English, however, was not completely satisfactory due to the lack of bilingual thesauri. In prior indexing work, the team at the UWM Libraries used a variety of controlled vocabulary tools in English, including TGM, (Thesaurus for Graphic Materials), TGN (Thesaurus of Geographic Names), AAT (Art and Architecture Thesaurus), and the Library of Congress Subject Headings. The Thesaurus for Graphic Materials was also used in this project to assign topical subject terms, but its application proved to be limited. There is no Chinese version of TGM and the selection of

English terms for the description of the visual content of Chinese scrolls and paintings was inadequate.

Library of Congress Name Authority File (LCNAF) was used to select authoritative forms of artists' names. The LCNAF lists both English and Chinese names and those forms were used when available. Not surprisingly, however, many names of the original creators are not listed in the LC Authority File. In those cases, the names were transcribed using the pronunciation approach based on Hanyu Pinyin and recorded in English following the established syntax of the LC Authority File.

The search for thesauri in Chinese demonstrated that there are very few similar tools available. AAT-Taiwan (Art & Architecture Thesaurus – Taiwan) represents a notable exception.

AAT-Taiwan, the Chinese version of the Getty's Art and Architecture Thesaurus, was developed as a result of the collaborative project between TELDAP and the Getty Institute (Chen et al., 2011). Getty's Art and Architecture Thesaurus provides access to English controlled vocabulary for indexing art, architecture, and objects of cultural heritage. AAT-Taiwan fulfills the gap for a similar vocabulary in Chinese. It proved to be extremely useful for subject term mapping in the *Tse-Tsung Chow Collection*. The terms for the Subject AAT field were selected from the English version, while the equivalent terms in Chinese came from AAT-Taiwan. Figure 4 demonstrates term pairs used for indexing a calligraphic couplet.

#### 5.4. Contextual Information

The original *Tse-Tsung Chow Collection* included very limited information about individual scrolls and paintings. In many cases, the people involved in the creation and custody of the original works, such as painters, calligraphers, and collectors, were not identified. The research process that was necessary to create bilingual metadata records and to provide access points in Chinese and English also resulted in a deeper understanding of the cultural significance of the described objects and the role of their creators. The collection includes 92 creators, and research conducted for this project allowed tracing and mapping the intricate relationships between painters, calligraphers, and collectors who often were poets themselves and powerful politicians. Biographical information was identified for 89 artists.

The importance of scholar-officials in the traditional Chinese social system emerges as an important theme in the research. The collection also reflects the political ideology and patriotism of scholar-officials in the late Qing dynasty. Over half of the identified authors are recognized as renowned politicians as well as scholars, writers and/or artists. Through the medium of traditional painting, they embedded their reactions to the political climate by making demands for greater political participation, or signaling their political disappointment and withdrawal from that world. One of artists featured in the collection is Qichao Liang (1873-1929), who was a renowned philosopher and political scholar in the late Qing dynasty (see Figure 1 and 4). Tse-Tsung Chow noted that Liang's work, Rengong Liang's (Qichao Liang) calligraphic couplet – 梁任公書集聯, was an expression of his frustration and pain after he lost his political mentor, Youwe Kang (1858-1927) and his friend Guowei Wang (1877-1927), a supporter of restoring the last emperor of the Qing dynasty. Kang and Liang were major leaders in the Hundred Days' Reform, a national reform movement that failed in 104 days in 1898.

## 6. Discussion

The project of creating bilingual metadata for the *Tse-Tsung Chow Collection* represents a practical solution to multilingual access in digital collections with a combination of human translation and vocabulary mapping. It also demonstrates that, in practice, multilingual access to collections of international origin may also require additional research to develop descriptive metadata as a foundation for translation and to provide the necessary cultural and historical context. Without additional contextual information, the meaning and cultural significance of digitized objects may not be readily apparent not only to English language users but also to Chinese language speakers who may not be familiar with traditional Chinese styles and scripts. The added value of transcribed and translated content extends the usefulness of the collection to students and scholars worldwide. However, the approach to multilingual access that involves research and relies on human translation is costly and requires language and subject expertise. It can only be undertaken in small scale-projects, like the *Tse-Tsung Chow Collection*, but is not feasible in large digital libraries. The automated translation methods, although imperfect, offer a greater potential for achieving multilingual access on a global scale, but it is worth remembering that machine translation needs to have a body of existing metadata and requires lexical augmentation or human review. As the test conducted for this project demonstrates (see Table 1), standard machine translation still fails to capture linguistic and cultural characteristics of heritage materials.

The promise of universal multilingual access in DLs has yet to be fulfilled. Rare collections of cultural and scholarly significance and those with unique international content may still require a “craftsman like” approach with human translation and professional indexing to capture the meaning and context of original works and to create records in the language and vocabulary of the target audience. Lor and Britz (2012) emphasize that the application of digital technology for digitization and preservation of cultural heritage implies obligations for holding and digitizing institutions. Although most of the authors’ discussion refers to the digitization of collections held in developing countries, the recommendation regarding the assignment of descriptive metadata is also pertinent to the collections of cultural heritage materials that originated in other countries. The authors stress that the use of descriptive vocabulary should reflect the language of the originating community (Lor and Britz, 2012). Multilingual access thus needs to be considered not only in the context of resource discovery but also as a way of connecting the digitized materials to their respective cultures and traditions.

## **7. Conclusion**

The project of building bilingual records for the *Tse-Tsung Chow Collection* at the UWM Libraries was undertaken as an experiment to expand access to the unique calligraphic scrolls and fan paintings to Chinese language speakers and users interested in Chinese cultural heritage. The process required extensive research and resulted not only in bilingual access to the collection, but also in enhanced descriptive records in both languages. The information gathered about the creators of the scrolls highlighted the importance of scholar-officials in the traditional Chinese social system. The effort invested in research and translation has not only contributed to bilingual access but also to enriching records in English. Multilingual indexing which involves extensive research, human translation, and vocabulary mapping is only possible in small projects. The approach is warranted in the indexing of unique collections of cultural heritage materials as it contributes not only to high quality descriptive records, but also to a deeper understanding of cultural context.



## References

- Borgman, C. L. (1997), "Multi-media, multi-cultural, and multi-lingual digital libraries: or how do we exchange data in 400 languages?", *D-Lib Magazine*, Vol.3 No. 6, available at: <http://www.dlib.org/dlib/june97/06borgman.html>
- Budzise-Weaver, T., Chen, J., & Mitchell, M. (2012), "Collaboration and crowdsourcing: The cases of multilingual digital libraries", *Electronic Library*, Vol. 30 No.2, pp. 220-232.
- Chen, J. and Bao, Y. (2009), "Information access across languages on the web: from search engines to digital libraries", 72nd ASIS&T Annual Conference, Vancouver, British Columbia, Canada, 6-11 November, American Society for Information Science and Technology (ASIS), Silver Spring, MD, available at: <http://www.asis.org/Conferences/AM09/open-proceedings/openpage.html>
- Chen, J., Azogu, O., Zhao, W., and Ruiz, M.E. (2011), "HeMT: A multilingual system for human evaluation of metadata records machine translation", in *Proceedings of the American Society for Information Science and Technology*, Vol. 48 No. 1, pp. 1-4.
- Chen, J., Ding, R., Jiang, S., & Knudson, R. (2012), "A preliminary evaluation of metadata records machine translation", *Electronic Library*, Vol. 30 No. 2, pp. 264-277.
- Chen, S., Cheng, C., & Chen, H. (2011), "Methodologies for multilingual information integration in the domain of Chinese art", *IFLA Journal*, Vol. 37 No.4, pp. 296-304, available at: [http://www.ifla.org/files/hq/publications/ifla-journal/ifla-journal-37-4\\_2011.pdf](http://www.ifla.org/files/hq/publications/ifla-journal/ifla-journal-37-4_2011.pdf)
- Chen, S., & Chen, H. (2012), "Mapping multilingual lexical semantics for knowledge organization systems", *Electronic Library*, Vol. 30 No. 2, pp. 278-294.
- Clough, P., & Eleta, I. (2010), "Investigating language skills and field of knowledge on multilingual information access in digital libraries", *International Journal of Digital Library Systems*, Vol. 1 No. 1, pp. 89-103.
- Diekema, A. R. (2012), "Multilinguality in digital libraries: A review", *Electronic Library*, Vol. 30 No.2, pp. 165-181.
- Gauthereau-Bryson, L., Estep, R., & Rivero, M. (2011), "Digitization practices for translations: Lessons learned from the Our Americas Archive Partnership Project", *D-Lib Magazine*, Vol.17 No. 9/10, available at: <http://www.dlib.org/dlib/september11/rivero/09rivero.html>
- Harpring, P. (2013), "Getty Vocabularies and linked data", pp. 1-46, available at: [http://www.getty.edu/research/tools/vocabularies/Linked\\_Data\\_Getty\\_Vocabularies.pdf](http://www.getty.edu/research/tools/vocabularies/Linked_Data_Getty_Vocabularies.pdf)
- Herold, I.H. (2010), "Digital archival image collections: Who are the users?", *Behavioral & Social Sciences Librarian*, Vol. 29 No.4, pp. 267-282.
- Jones, G. J., Zhang, Y., Newman, E., Fantino, F., & Debole, F. (2007), "Multilingual search for cultural heritage archives via combining multiple translation resources", in *Proceedings of the Workshop on Language Technology for Cultural Heritage Data*, Prague, June 28, Association for Computational Linguistics, Stroudsburg, PA, pp. 81-88, available at: <http://acl.ldc.upenn.edu/W/W07/W07-09.pdf#page=91>
- Kraus, R. C. (1991), *Brushes with Power: Modern Politics and the Chinese Art of Calligraphy*, University of California Press, Berkeley.
- Kupietzky, A.S.G. (2007), *Subject Access to a Multilingual Museum Data Database: A Step-by-Step Approach to the Digitization Process*, Libraries Unlimited, Westport, CN.

- Lor, P.J. & Britz, J.J. (2012), "An ethical perspective on political-economic issues in the long-term preservation of digital heritage", *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 11, pp. 2153-2164.
- Ménard, E. (2010), "Ordinary image retrieval in a multilingual context: A comparison of two indexing vocabularies", *Aslib Proceedings*, Vol. 62 No.4/5, pp. 428-437.
- Ménard, E. (2012), "TIARA: The 'making of' a bilingual taxonomy for retrieval of digital images", *Library Hi Tech*, Vol.30 No.4, pp. 643-654.
- Van Oudenaren, J. (2010), "Connecting the world, responding to user needs", *Information Outlook*, Vol.14, pp. 10-12.
- Wu, D, He, D, & Luo, B., 2012, "Multilingual needs and expectations in digital libraries: A survey of academic users with different languages", *Electronic Library*, Vol. 30 No.2, pp. 182-197.