



The Multilingual Full-text Database of Overseas Life-writing on Modern Chinese People and the China Biographical Database

Rong Huang

Peking University

The Multilingual Full-text Database of Overseas Life-writing on Modern Chinese People

<http://sclw.lib.sjtu.edu.cn/index>

The Multilingual Full-Text Database of Overseas Life Writing on Modern Chinese People (the Database) was launched by the Centre for Life Writing at Shanghai Jiao Tong University (SJTU) in November 2011. All subjects in the Database are influential, representative, or exemplary Chinese people whose life materials are either published or preserved outside mainland China after 1898, the year generally considered to be the beginning of China's Modern Age. The Database aims to bring a prosopographical perspective to the transnational and transcultural experience shared by a rather large number of Chinese people in the previous century. By April 2018, the Database has listed more than 3,000 subjects and over 15,000 relevant resources.

In many ways the Database is the first of its kind. To start with, the subject matter is highly significant. According to the Database, more than 15 million Chinese people went overseas during this period. Together they form a formidable group which has profoundly transformed the world in the 20th Century, and whose influence and achievements are still widely acknowledged today. The Database's 3,000 subjects are elites from this group. They are divided according to their professions into 11 categories, which point to a strong relationship between career and identity, corresponding with the time when modern divisions of labour started to take shape in China.

Geographical inclusiveness is another prominent feature. The Database consists of ten sub-projects, nine of which are dedicated to the

collection of life materials in different areas of the world, whilst the remaining one is exclusively for the construction of the Database's digital framework. Each of the nine sub-projects has one centre as its geographical base to collect and arrange materials from the area; the nine centres are the U.S., the U.K., France, Germany, Russia, Japan, Singapore, Taiwan, and Hong Kong and Macao. These are the destinations that used to attract the most Chinese people and which therefore now hold the largest quantity of their life materials. Geographical visualisation reveals interesting trends in the distribution of population, with Europe being one of the most favourite destinations. The four centres in Europe contribute almost a third of the subjects and a total of 1,563 resources. The nine centres speak for the special experience of modern Chinese diasporas.

Moreover, life materials are presented by the Database to be multi-generic, multimedia and multilingual. Primary sources include autobiographies, biographies, letters, diaries, memoirs, travelogues, dictionaries, interviews, newspaper reports, periodical coverage, archival miscellanea, oral history accounts, photographs, recordings, videos, and online entries in reference works. The variety of genres suggests an understanding of life writing in its widest possible sense. The Database also covers primary materials in six languages, including Chinese, English, French, German, Russian, and Japanese, which ensures an international reach for the entire project. For each search result, there is a link to the full-content, copyright-protected text. As to content that is not available in the full text, there will be necessary information—for instance, book covers, publishing details or the table of contents—to direct the user to the right resource. The Database thus makes itself an invaluable archival hub for users who hold a general interest in modern Chinese history and life writing.

An example helps to show how the database works. Consider the case of Hu Shi (1891–1962), a major spokesperson for China's New Cultural Movement during the early 20th Century. Enter his name in Chinese characters and there will be 149 results. Each result is displayed by title, author, language, publication date, and country of origin. Results can be filtered accordingly. If the user wants to see the full text of a certain result, a new window will be opened to redirect the user to a copyright-protected online reader, where the content can be accessed. The scope of the resources, the effortlessness of the search, and the accessibility of the full-text content make the Database a favourite starting point for scholars of not only modern Chinese life-writing studies but also modern Chinese history.

Even for such a powerful Database, language, not surprisingly, turns out to be a problem, especially for this period in the Chinese history.

Hu Shi has at least two kinds of spellings for his name. Apart from “Hu Shi”, as in *pinyin*, the standard Romanization system for the Standard Mandarin Chinese today, his name also appeared as “Hu Shih” in the Wade-Giles Romanization system that was commonly used during the first half of the 20th Century. Search for “Hu Shi” in the Database and there are 5 results; search for “Hu Shih” and there are 25. Keeping different spelling systems in mind is of vital importance because for the first time in history, Chinese people began to make their voices heard in Romanised languages that were not native to them. Hu Shi himself delivered speeches and published articles in English, so as to reach an international audience. It is therefore crucial to know how his name is romanised in English so that no life materials are missed from the search.

In 2016 the Database completed its first-phase construction. Academic support and collaboration were the two keys to its success. The Database was designed as a part of the Major Project “Compilation and Research of Overseas Life Writing on Modern Chinese People,” sponsored by the China National Social Sciences Fund. The other major sponsor is SJTU, which allocated funds with amounts that equalled those provided by the China National Social Sciences Fund. Over 20 scholars in life writing and digital humanities from all over China and abroad contributed to the construction of the Database. Institutions, libraries, archives, media branches, scholars, and private collectors around the world kindly offered their help. The Database is indeed an outcome of collective efforts.

Here is a final note about some promising fruits derived from the Database. Today digitization has developed from a cutting-edge trend into a sought-after reality. Going digital appears to be imperative. That said, it remains premature to announce the death of the text. Much like the way that life-writing scholars in this Information Age benefit from traditional training based on the reading of texts, the Database also values textual output. Hence a ten-volume publication incorporates current research findings related to the Database. Each volume concentrates on research results from one geographical centre and the final volume is compiled as the index of the full Database. Monographs are produced, too. Among them, *The Management and Construction of Life-Writing Resources* is the first book in China and possibly in the world that foregrounds the combination of life-writing and information technology. Articles about the Database are regularly published in the *Journal of Modern Life Writing Studies*, the first Chinese journal that is exclusively dedicated to life writing. The Database has evolved into a platform where online and offline applications go hand in hand.



Screengrab 1: Multilingual Full-text Database of Overseas Life-writing on Modern Chinese People, Homepage, accessed 20 April 2018

CHINA BIOGRAPHICAL DATABASE

<https://projects.iq.harvard.edu/cbdb>

The China Biographical Database (CBDB) is a large data aggregator and powerful content harbour co-developed by Harvard University, Academia Sinica, and Peking University. It is one of the oldest and biggest digital humanities projects focusing on China. In its latest data release in August 2017, CBDB contains biographical information about 417,000 Chinese individuals, mostly from the 7th century to the 19th century. CBDB looks for materials beyond the lives and careers of historical individuals; it incorporates specific interactions among people, so that the interpretation of a person's life is not limited just to his or her story but extends to including the person's role in an entire community.

Unlike many text-based databases, CBDB does not preserve primary life materials in the original textual sources. Rather, it mines texts from digital sources of reliable historical records and stores them in a relational database. Current sources include modern syntheses of biographical data (e.g. biographical indexes for different historical periods), traditional biographical records (mainly biographies in China's Standard Histories), evidence for social associations from literary collections, evidence for office holding from modern and traditional sources, and relevant biographical databases developed in China, Taiwan, Japan, and other regions. This approach guarantees efficiency and enables CBDB to accumulate data at an unprecedented speed compared to manual input.

For data mining to be successful, the processing of key attributes in biographies is crucial. CBDB assigns three major entities: people, offices, and association types. Individuals are identified by three groups of information: their basic biographical information (birth name, alternate names, and places of birth and living, etc.); their professional experience (writings, postings, and modes of entry into government); and their social relations (kinship, other forms of associations, and social status). Relevant data on events are also included if necessary. Such information helps to locate the individual in groups of people socially, temporally, and geographically.

The value of CBDB's data is not only in the basic biographical information of individuals. Search in CBDB's online query system for the name "Zhu Xi," the most influential philosopher during the Song Dynasty (960–1279), for example. The database yields three results: Zhu, his mother, and his wife. Each of the entries is displayed in attributes including result number, name, dynasty, place of birth, tribe, alternate names, foreign names, and index year (the year when the person reached the age of 60 or the year of death if the person demised before 60, provided that such information exists). Click on the link of Zhu Xi, and the user will be directed to a more detailed set of information arranged according to the identifying attributes. All results are shown in tables that contain the necessary information on the online system. Such descriptions of biographical data often prove to be daunting for scholars who intend to conduct thorough research on specific historical figures. This, nevertheless, does not lead to the conclusion that CBDB is of no use for life-writing researchers. Based on an Internet search done on 20 April 2018, CBDB reveals that Zhu held 32 different kinds of government titles in his lifetime and that 2,566 people were associated with him. Numbers like these and their specific records could bring new light on existing life-writing studies.

CBDB's real strength lies in its huge potential in prosopographical studies. The amount of data in mature databases like CBDB resists close reading of entries one by one; visualisation of large sets of data becomes an advantageous method to demonstrate demographic and other patterns. Thus CBDB poses new challenges and also excellent opportunities for life-writing scholarship, which is usually based on close-up textual analysis. Prosopography asks questions not merely about the collection of data, but more importantly about aggregate data on social groups.

The CBDB project has been in existence for almost 15 years, but its data is underutilised by humanities research. The main reason is the insufficient comprehension of data among humanities scholars, and the fact that many have yet to realize the potential of data technologies in the humanities. For sure, things are beginning to change, especially due to

recent interdisciplinary collaborations within and outside China. Digital Humanities, for example, has gradually been established as a proper field of research all over the world. How life-writing scholars approach its latest developments will be a challenge.

While the subject matter may attract only specialists in Chinese history, the overall vision of CBDB and its application in biographical studies provide exemplary references for life writing in general. Generations of designers have actively incorporated the latest technological developments into the database design. Detailed documentation on its design and usage is available on the website. Materials from training programmes, presentations, data sets, and supporting software including the database as a standalone file and as an API interface are all accessible there. Today, scholars from history, life writing, and digital humanities are engaging CBDB in their projects to present previously unnoticed traces of people and the histories they have created.

One of the latest developments of the database (in March 2018) is that the CBDB project has signed a cooperation agreement with the ChineseAll Digital Publishing Group Co., Ltd., the first digital publishing stock in China. Again CBDB has made itself a forerunner in promoting digital humanities and the commercialization of historical data in China. Both



China Biographical Database Project (CBDB)

HOME About Us Methodology Sources Events Download Blog 中文

Welcome

Introduction

The China Biographical Database is a freely accessible relational database with biographical information about approximately **417,000 individuals as of August 2017**, primarily from the 7th through 19th centuries. With both online and offline versions, the data is meant to be useful for statistical, social network, and spatial analysis as well as serving as a kind of biographical reference. The image below shows the **spatial distribution** of a cross dynastic subset of 190,000 people in CBDB by basic affiliations (籍貫).

How to Cite CBDB

[Harvard University, Academia Sinica, and Peking University, China Biographical Database \(January 1, 2018\). <https://projects.io.harvard.edu/cbdb>.](https://projects.io.harvard.edu/cbdb)

Updates

(April 2017) **CBDB's 2017 April update is released!** Download the standalone database [here!](#)

(April 2017) **New blog post about ChinaX is up!** See [here](#).

(April 2017) **Ming Government Official Titles: A Crowd-Translation Project**

Screengrab 2: China Biographical Database, Homepage, accessed 20 April 2018

parties consider the cooperation a win-win opportunity, in which the ChineseAll will provide funds and technical support to ensure the sustainability of CBDB, while gaining from the gigantic data, the time-tested framework, and the renowned academic reputation of the database project. The challenge lies in how this collaboration balances academic and commercial needs. ChineseAll is experienced in content production for the general public, whereas CBDB is established as an enterprising project for incubating cutting-edge academic research on China. How the two are going to make use of each other's strengths will be the key to the future of the use of biographical data from China's history.

ACKNOWLEDGEMENTS

For the Database part, I must thank those in charge of the overall planning, coordination and operation of the project including Chief Expert Yang Zhengrun, Academic Secretary Liu Jialin, and Academic Assistant Yuan Qi for their late-night reviews of my writing. For the CBDB part, I am grateful to Dr. Lik Hang Tsui from the China Biographical Database Project for his historian's precision and insider's expertise. My thanks also go to Dr Rob Gallagher, Professor Clare Brant, and the editors of the European Journal of Life Writing for their generous suggestions.

ABOUT THE AUTHOR

Rong Huang is currently a postdoctoral fellow at The School of Health Humanities, Peking University, where she works on the intertwining between life writing and medical humanities. She holds an MA in English Language and Literature and a Ph.D. in Comparative and World Literature. Her research interests include autobiographical narratives, contemporary world literature in English, theories of the body, and media studies. She has published articles about life writing, world literature and new media. Email: huangrong@pku.edu.cn.