INTRODUCTION: CHINESE STUDIES AND CHARACTER ENCODING

It was during the 1980s in the midst of the trend towards computer downsizing that scholars and administrators in the fields of the Humanities and Social Sciences came to the realization that the introduction of the personal computer (PC) had become as necessary as the bookshelves in their libraries. However, this PC revolution failed to bring about immediate innovations to help the discipline of Chinese studies develop in a significant way, information-wise. The main reason why was the difficulty in using ideographs like kanji, hiragana and katakana with the processing technology available at the time. Ideograph processing experts had to begin at the first stroke of the first kanji character and create huge character-by-character fonts, then figure out how they were going to be used. In Japan, after this challenge produced four protocols, GT 書体, e 漢字, JIS 第三 and 第四水準 and packages like 超漢字 (180,000 chrs.) and 今昔文字鏡 (150,000 chrs.) began reaching end users, the first stage in the digitalization of Chinese studies was accomplished [Nikaido 2001]. However, at the present time, politics is rearing its ugly head in the face of a battle for world supremacy between
Unicode 3.1 and the People’s Republic of China’s GB18030, which is trying to absorb Unicode. No matter who’s side one may be on, the appearance of Windows XP and Mac OS X have eliminated the need to purchase Chinese input software; and such simplified input methods as Pentablet and voice recognition have significantly reduced stress both on user pocketbooks and nerves.

Nevertheless, the quest for a penultimate kanji code is far from over. Actually, conventional character code designers have only paid attention to the glyph (topological) aspect, while ignoring other aspects such as pronunciation, meaning and code information. (This has been pointed out by Kyoto University’s CHISE (See Database Listing (hereafter, DB) at the end of this article.)) Besides encoding, interest is being drawn to such diverse aspects of original kanji forms, historical variations and personal calligraphy styles. For example, the Kyoto University Institute of Humanities is now working on a system for searching kanji as ideographs for such sources as the ancient rubbings and tortoise shell paleographs it holds (DB). It is in this way that when the challenge of character encoding was first taken on during the 1980s, the resulting databases emphasized bibliographic and subject searching. It was a time when the Internet was still in its infancy and the database itself was all but ignored as a device for sharing research data internationally. For this reason, the use of databases spread on a “stand-alone” basis, and even at research institutes sharing huge computers, information circulated on a strictly intramural (intranet) basis.

I. FROM PRINTED CATALOGUES TO DIGITAL DATABASES

During the 1990s, when the World Wide Web was developed and the Internet became commercialized, attention was drawn to the possibilities of their application to Chinese studies. In Japan, particularly around 2000, epoch-making attempts were begun to convert bibliographic data, which had theretofore been disseminated through the print medium, into the digital medium. One of these attempts involved Chinese bibliographical studies. In 1999, it was decided that the Kyoto University Institute of Humanities’ Asian Bibliography Center would be reorganized as the Kanji Information Research Center beginning in April of the following year, while the University of Tokyo Institute of Asian Culture’s Information Center was set up along with the National Information Sciences Institute and National Chinese Bibliographic Database Association. Since many Chinese sources had not yet been reg-
istered in the Online Public Access Catalog, the printed catalogs from all the Asian research libraries, including the Toyo Bunko, Institute of Humanities and Institute of Asian Culture had to be employed. After lugging around literally tons of volumes containing information on the whereabouts of Chinese holdings in Japan, the National Chinese Bibliographic Database was compiled (DB\textsuperscript{3}) in 2001, together with the appearance of the Toyo Bunko holdings online (DB\textsuperscript{4}). The year 2000 marked the beginning of attempts to utilize online 『東洋学文獻類目』Asian Studies Bibliography, a series that had been published by the Institute of Humanities since 1934 (DB\textsuperscript{2}).

One more model tested during this time was “Scripta Sinica” at the Academia Sinica Computing Center in Taiwan, which today offers about 50 million characters of complete text related to Chinese history, the classics and literature. Complete text searching is different from the conventional bibliographic and subject index databases in that source materials themselves can be handled on a computer screen. Users all over the world now understand the significance of being able to use alliances of character fonts on the same level when dealing with source materials, since they now realize that the existence of only bibliographic databases still necessitate having to go to the repositories to view the actual content, resulting in no appreciable improvement in utilization efficiency. Japanese users have been made to see things this way by the Institute of Humanities Central Asia Travelogue Database (DB\textsuperscript{2}) and the Niigata University’s Chinese Classical Database (DB\textsuperscript{6}), both complete text websites. Setting up complete text search engines requires huge amounts of both labor and time, much of which in recent years both Taiwan and the People’s Republic have generated, in comparison to Japan’s smaller scale, individually based efforts.

Two typical examples are 1) the marketing of a 800 million text and pixel search engine for the Siku Quanshu 四庫全書, dynastic China’s largest and most comprehensive compilation project, requiring thirty-one years to complete (1741–1772) and 2) the 1 million selection (2 billion characters of text) AncientBookClient software. In contrast, Japan has gradually moved away from complete text database creation to collections of image facsimiles. The best example of this is the Japan Center for Asian Historical Records (JACAR) which was opened in the National Archives during November 2001 and now makes available on its website source materials held by the Archives, the Foreign Ministry Diplomatic History Museum and the National Defense Research Institute Library (DB\textsuperscript{7}). JACAR has also stimulated archives and libraries in Korea and Taiwan to make their holdings available interna-
tionally, thus helping to put researchers throughout the region and the world “on the same page.” One more similar project is the National Diet Library’s Modern Digital Library containing its uncopyrighted holdings from the Meiji and Taisho eras (DB⁸), offered free of charge. Similar efforts by universities are now underway with the Institute of Humanities “Asian Studies Digital Library” (DB⁹), the Institute of Asian Culture’s “Rare Chinese Books Image Collection” (DB¹⁰) and Otaru University of Commerce’s “Rare Books Image Data: China” (DB¹¹). There is also the unique digital newspaper collection at Kobe University featuring economic-related articles from prewar Japan (DB¹²), boasting around 500,000 items in either image or text form, many touching upon Sino-Japanese international relations. The JETRO Institute of Developing Economies database, “Japan in Modern and Contemporary Asia” has been made available to the public, featuring about 21,000 monographs and yearbooks and 1243 periodical titles related to the former Japanese colonies of Taiwan, Korea and Manchuria. In toto, about 120,000 volumes are available for search on Japan, the United States and China (DB¹³). The project, which focuses on the study of colonialism, is expanding to include data shared with the Toyo Bunko.

II. COMPOSITION OF THE IMAGE DATABASE

The attention which has been drawn towards non-textual source materials in postwar Japan created a “cultural turn” during the 1980s, leading to a reevaluation of the field of cultural history. The publication by Heibonsha Limited, Publisher between 1986 and 1994 of its Image Reading Series is one concrete example of this phenomenon, especially the title, Wargraphy: The People Who Created “The Front.” The means to take such non-textual image data off the printed page and put it into cyberspace was accomplished by the creation of the high resolution JPEG image compression format in 1994 and the development of all the digital cameras, scanners and monitors able bring it onto our desktops. Efforts have continued since 2003 at the Toyo Bunko, beginning with the “Digital Silk Road Project” combining information science with cultural heritage, resulting in its present multi-media library (DB¹⁴). At the same time, the Toyo Bunko has added an image database showing users the unique holdings its Library has in store for readers, featuring woodblock prints, watercolors, cartography and pamphlets that were collected from the time that Iwasaki Hisaya purchased in 1917 the personal
library of G. E. Morrison, the veteran journalist and diplomat to late Qing and early Republican China (DB4).

Hitotsubashi University’s photograph collection of pre-WWII Asia, features nearly 5000 prints collected by the Army University of Economics from such areas as central and north China, Manchuria and Mongolia (DB3). There is also the “Thomson Photograph Collection of 19th Century China” at the Tokyo University of Foreign Studies (DB14), focusing on John Thomson’s (1937–1921) oversized publication, Illustrations of China and Its People (London, 1873–1874), featuring 200 photos dealing mainly with daily life in China under the Qing Dynasty circa 1870. An upcoming, and very exciting, project is the digitalization of the over 70,000 news photographs related to Manchuria and China now stored in the Asahi Shinbun Osaka Headquarters’ Fuji Warehouse.4

III. REGIONAL STUDIES AND INFORMATION SCIENCE: CROSS-SEARCHING

In the midst of a growing interest in regional studies, attempts as database creation in China research began from the 1990s on to reexamine methods for emphasizing personal names, historical eras and keywords, in order to focus on specific geographical areas crisscrossing with and expanding into the world at large. Consequently, we were confronted with the issue of how to develop an approach integrating the three axes of people, time and space.

As to what this meant for website databases, it was found that the inconvenience involved in having to use individual bibliographical databases scattered all over the globe could be overcome by either search engines that traversed them or by resource sharing. Consequently, the call went out for both integrated source material searching and the creation of databases where original sources could be viewed regardless of location or research situation. The database attempting to prepare a research platform around the theme of “Japan and the World” being developed by Prof. Tanaka Akihiko’s team at the Institute of Asian Culture is one case in point (DB9). This valuable DB includes a chronology spanning 1920–1995, topics related to postwar Japanese and Chinese domestic politics and the foreign affairs of all countries in the Asian-Pacific region. One part of the Tohoku University NOAH Image DB (DB15) includes a section on China created from view and original satellite images of a geographical expanse of some 1100km from...
Shanghai to Beijing onto Heilongjiang Province. In sum, the time when “China” was the exclusive bailiwick of the Humanities is now over, as more and more academic disciplines are making Sino-commitments in their research activities and consequently demanding databases that will serve their specific needs.

One example of this trend is DBᶒ, a bibliographic database for doing social science-related research on Northeast Asia, which this author made available by JSPS for Grants-in-Aid for Scientific Research in March 2004. This particular DB allows cross-searches of “Chronological Table of 20th Century: 1918–1952,” “Municipal Government Bulletin of Beijing Special Municipal, Shizheng Gongbao 市政公報 (1938–44),” “Shanghai Municipal Police Files (1894–1949),” “China-related material archives, Hoover Institute, Stanford University” and “Subject Index of Humanities and Social Science Periodicals published by the Mongolian Academy of Sciences” [Kishi, Ishikawa, and Inoue 2004]. One more attempt at cross-searching is the portal site, “War and the Media” which this author also started in conjunction with several research institutes and libraries (three at present) to integrate their databases of non-textual sources. We look forward to welcoming more participants in the future. At present, the portal site allows users to search a propaganda poster collection from the World War I era, a DB of picture postcards from pre-WWII East Asia [Kishi 2006], an East Asian media image DB, and posters and bills related to the topic, “Manchuria and the Media.” Soon a source collection of propaganda circulated by the Japanese Cabinet Public Relations Section will be made available.

IV. USING THE GEOGRAPHICAL INFORMATION SYSTEM

In the aftermath of the Great Hanshin Earthquake of January 1995, the Japanese government went to work popularizing the geographical information system (GIS), adopting in December of the following year a long term plan to build a national topographical database and promote the spread of GIS technology. 1995 was also the year that the Japanese version of the ArcView GIS software was first released. Today researchers are not at all satisfied with databases, be they bibliographic, textual or image, that are already “precooked,” preferring those that can be altered and orchestrated to their particular needs, as long as the data is reliable. This an attitude that must be kept in mind when trying to accommodate information science to Chinese studies. Experimenting with GIS and
remote sensing is a case in point, not only in studying China, but also doing regional studies in general.

Any discussion of this subject would never be sufficient without mentioning UC Berkeley’s Electric Cultural Atlas Initiative (ECAI) or Taiwan’s Academia Sinica’s Pacific Neighborhood Consortium (PNC).\(^5\) ECAI involved the design of an application for expressing cartographic information on the Calmap system website, while PNC developed a full text DB covering the Institute’s output. In addition, the fruits of these two projects resulted in cooperation between Harvard and Shanghai Fudan University in producing a China Historical GIS (DB\(^8\)). ECAI was also inspirational in Sidney University of Archeology Professor Ian Johnson’s development of the exciting TimeMap software.\(^6\)

In Japan, it was in 1989 that under the auspices of the Humanities and Computer Science Research Association (Jinmonkon) took the lead in creating a GIS database and popularizing the technology. Today one of the most active groups in building a field of regional information science is the H-GIS Research Association led by Shibayama Mamoru, professor at Center for Southeast Asian Studies, Kyoto University and Hara Shoichiro, professor at that University’s Center for Integrated Area Studies.

The Association, which dates back to 2004, employed GIS technology based on Johnson’s TimeMap in a user-friendly Japanese language environment to come up with HuMap, a time and space mapping tool now being tested by the inter-university National Institutes for the Humanities (NIHU). The application of this system’s region-information linkage to Chinese studies has been studied, first at the April 2005 ECAI Shanghai Conference’s Panel 5: Special History and Visual Documents, then at the February 2007 Regional Studies and Information Science Conference held at Kyoto University under the theme “Opening New Horizons.” The former gathering was an attempt to confirm through concrete examples the applicability of digitalization to the study of China, while the latter, although not a Chinese studies venue per se, brought regional studies and information science experts together to discuss the most recent developments of joint interest.\(^7\) The International Japanese Cultural Research Center devoted its 24\(^{th}\) Research Conference in February 2005 to the theme, “Reading Historical Space: The Application of GIS to the Study of Culture and Civilization” [Uno 2006]; and during August 2007, Nagoya was the sight of a symposium on GIS and historical maps.\(^8\) Such gatherings imply that those in Japan genuinely dedicated to finding GIS applications for Chinese studies find themselves mostly involved in geographical and
archaeological sciences, but also patiently await participation from other disciplines as well.

V. THE RAMIFICATIONS FOR CHINASTUDIES

As China’s presence in world events increases, it will not only continue to be studied in its historical context, but also as an important key to analyzing contemporary issues. One important approach to contemporary Chinese studies is its articulation with information science. The idea is that research and source materials should not be limited to individuals or groups, but rather shared among all of them in the interest of advancing the world’s knowledge about China. For example, Chinese departments at Leiden and Heidelberg Universities have since 1995 been operating a website entitled “Internet Guide for Chinese Studies” with funding from the Berlin National Library and the German Research Fund (DB19). Another similar attempt has been made by Asian studies scholars at the University of Tokyo in the form of the Asian Research Information Gateway, an open forum offering guides to libraries and archives, and reporting on related seminars and lectures being held both in Japan and abroad (DB20). The aim is not only to accumulate a substantive research database, but also build a platform for scientific investigation into the field of Chinese studies.

Despite all of this forward-looking activity, Japan has yet to see a full-blown, China research information center open to the public. It in this sense that we look to the interuniversity Human Culture Research Organization’s present contemporary China program, involving six ongoing projects. Within this joint effort, the Toyo Bunko’s Contemporary China Studies Archive will be playing a very important role in terms of data accumulation and dissemination.9

Digitalized information and its public availability has become an important issue not only in China studies, but also in every aspect of contemporary Japanese life, as evidenced by the huge amount of commercial databases flooding the market these days. On the other hand, we in China studies are faced with scattered, isolated DBs that need to be linked and portalized and are still burdened with conventional geographical data that needs to be digitalized in some way, shape or form. The challenge facing us in the form of joint research today transcends the organizational limits of single research institutes, forcing us to cross national borders and get involved in the kinds of projects pioneered by ECAI and PNC, mentioned above. The promotion of internationally
linked projects is one of the quickest avenues for connecting Chinese studies to information science, and vice versa. It is that kind of mindset that should form the academic standard for creating databases, for if not, we are in serious danger of being inundated and swallowed up by the huge sea of information that has flowed out of modern and contemporary China.

—*Originally written in Japanese*

### NOTES

1. The revised edition is available at [http://www2.ipcku.kansai-u.ac.jp/~nikaido/uni_kanji.html](http://www2.ipcku.kansai-u.ac.jp/~nikaido/uni_kanji.html).
2. At the present time in the kanji use sphere, there are the Chinese GB code, Taiwanese Big5 code, Japanese JIS code, Korean KS code, and European HZ code; but there is also the CJK code, which integrates all the above into Unicode and, which, for better or worse, opens the way to a solution, exemplified by the Mojikyo international Unicode font set.
3. The best reference available at this date on computing with kanji can be found in the journal of the Kanji Bunken Johoshori Kenkyukai 漢字文献情報処理研究会 (Japan Associate for East Asian Text Processing), *Kanji Bunken Johoshori Kenkyu* 漢字文献情報処理研究 (Journal of Japan Association for East Asian Text Processing) and its e-mail magazine.
4. An introduction to what is available appears on the cover design of *Asahi Shinbun Shuzaihan 2007*.
5. ECAI’s website is [http://www.ecai.org/](http://www.ecai.org/); PNC’s [http://pnclink.org/](http://pnclink.org/). This author was able to participate in a joint session involving both projects held in Berkeley during form 18 to 20 October 2007, entitled “Area Studies, Then and Now.”
7. The program and report summaries of the ECAI Shanghai Conference can be found at [http://www.ecai.org/Activities/shanghai2005/ECAL_Print.doc](http://www.ecai.org/Activities/shanghai2005/ECAL_Print.doc). For the proceedings of the latter, see [Shibayama et al. 2007].
8. The proceedings are available in [Mizoguchi and Ishikawa 2007], but unfortunately do not include the paper given by G.W. Skinner.
SELECTED DATABASE LISTING (DB)

① CHISE (CHaracter Information Service Environment)／守岡知彦
Morioka Tomohiko (Kyoto University)
http://mousai.kanji.zinbun.kyoto-u.ac.jp/ids-find

② 京都大学人文科学研究科 Institute for Research in Humanities, Kyoto
University
東洋学文獻類目検索 Bibliography of Oriental Studies on the Web
(CHINA3), ver. 6α
http://mousai.kanji.zinbun.kyoto-u.ac.jp/ruimoku/index.html.ja
東方学デジタル図書館 Asian Studies Digital Library
http://kanji.zinbun.kyoto-u.ac.jp/db-machine/toho/html/top.html
石割拓本資料 Stone Rubbings in the Holdings of the Institute for
Research in Humanities
http://kanji.zinbun.kyoto-u.ac.jp/db-machine/imagesrv/takuhon/
甲骨文字 Oracle Bone Script http://mousai.kanji.zinbun.kyoto-
u.ac.jp/koukotsu/
西域記データベース Database of Texts and Photographs focusing
on Central Asia http://kanji.zinbun.kyoto-u.ac.jp/~saiiki/

③ 全国漢籍データベース A Database of the Chinese Classics in Japan／
全国漢籍データベース協議会 Japan Society of the Chinese Classics
http://kanji.zinbun.kyoto-u.ac.jp/kanseki/

④ 東洋文庫 The Toyo Bunko (Oriental Library)
東洋文庫所蔵漢籍オンライン検索 The Toyo Bunko Online Classical
Chinese Title Search
http://61.197.194.10/TBDB/KansekiQuery3.html
東洋文庫所蔵図像史料マルチメディアデータベース文庫 Digital Archive
of Toyo Bunko Rare Books
http://dsr.nii.ac.jp/toyobunko/
画像データベース Image Database
http://www3.toyo-bunko.or.jp/gazou/gazou_index.html

⑤ 漢籍電子文献 Scripta Sinica／台湾・中央研究院 Taiwan Academia
Sinica
ver.1 http://www.sinica.edu.tw/ftms-bin/ftmsbw3/
ver.2 http://www.sinica.edu.tw/~tdbproj/handy1/
ver.3 http://140.109.138.237:9080/Handy/

⑥ 中国思想史研究室の書庫 Department Library of Chinese History of
Ideas／新潟大学人文学部紀玉憲明研究室 Kodama Noriaki’s Office,
Faculty of Humanities, Niigata University

⑦ アジア歴史資料センター Japan Center for Asian Historical Records／国
立公文書館 National Archives of Japan
http://www.jacar.go.jp/
⑧近代デジタルライブラリー Digital Modern Library／国会図書館
National Diet Library
http://kindai.ndl.go.jp/index.html
⑨東京大学東洋文化研究所 The Institute of Oriental Culture, The
University of Tokyo
貴重漢籍善本全文画像 Full-text Database of Chinese Rare Book
http://shanben.ioc.u-tokyo.ac.jp/
東洋文化研究所所蔵漢籍目録 Database for the Catalogue of Chinese
Book Belonging to the Institute
http://www3.ioc.u-tokyo.ac.jp/kandb.html
世界と日本 The World and Japan／田中明彦研究室 Tanaka Akihiko’s
Office
http://www.ioc.u-tokyo.ac.jp/~worldjpn/index.html
⑩貴重図書全文画像データ(漢籍) Full-text Database of Chinese Rare Book
／小樽商科大学附属図書館 Otaru University of Commerce Library
http://www.otaru-uc.ac.jp/htosyo1/siryo/kanseki/
⑪神戸大学戦前期新聞経済記事文庫 Digital Archive for Newspaper
Article, Kobe University Libraty／神戸大学経済経営研究所図書室
Library of Research Institute for Economics and Business
Administration, Kobe University
http://www.lib.kobe-u.ac.jp/sinbun/index.html
⑫近現代アジアのなかの日本 Japan in Modern Asia／ジェトロ・アジア経済
研究所 Asian Research Center, IDE-JETRO
http://opac.ide.go.jp/asia_archive/index.html
⑬戦前期アジア諸国写真コレクション Photo Collection of Pre-1945 Asian
Countries／一橋大学機関リポジトリ Hitotsubashi Educational and
Research MEdia Service-IR
http://hermes-ir.lib.hit-u.ac.jp/da/handle/123456789/18
⑭トムソン写真の世界／19世紀中国の人々 John Thomson’s Photo World
—Chinese People in 19 Century／東京外国語大学アジア・アフリカ言
語文化研究所 ILCAA, Tokyo University of Foreign Studies
http://irc.aa.tufs.ac.jp/thomson/top.html
⑮東北大学ノア画像データベース（中国）Tohoku University NOAA
Image Database (China)／東北大学東北アジア研究センター CNEAS,
Tohoku University
http://asiadb.cneas.tohoku.ac.jp/china/
⑯北東アジア地域の社会科学研究のための資料・書誌情報データベース A
Database of Documents and Bibliographies for Social Sciences in
Northeast Asia／貴志俊彦 KISHI Toshihiko (Kanagawa University)
ポートサイト「戦争とメディア」Portal Site “Words and Images of War —Inter-Institute Archive of 20th Century Media Collections”／吉見俊哉・貴志俊彦・孫安石 YOSHIMI Shunya, KISHI Toshihiko, SON An-Suk
http://media-portal.iii.u-tokyo.ac.jp/

China Historical GIS／Center for Geographical Analysis, Harvard University ＆ 復旦大学歴史地理研究中心 The Institute of Historical and Geographical Research, Fudan University English Version
http://www.fas.harvard.edu/~chgis/
Chinese Version http://yugong.fudan.edu.cn/Ichg/Chgis_index.asp

Internet Guide for Chinese Studies／Sinological Institute, Leiden University, Netherlands
http://www.sino.uni-heidelberg.de/igcs/

アジア研究情報ゲートウェイ Gateway to Asian Studies in Japan／東京大学The University of Tokyo
http://asj.ioc.u-tokyo.ac.jp/html/information.html

SELECT BIBLIOGRAPHY

Asahi Shinbun Shuzaihan 朝日新聞取材班, ed. 2007. 『過去の克服』と爱国心 (Patriotism and “getting over the past”)．Tokyo: The Asahi Shimbun Company (朝日新聞社)．
Shibayama Mamoru 柴山守, et al., eds. 2007. 「シンポジウム地域研究と情報
学：新たな地平を拓く」(Symposium on regional studies and information science: Opening new horizons). Kyoto: Center for Integrated Area Studies, Kyoto University (京都大学地域研究統合情報センター), Center for Southeast Asian Studies, Kyoto University (京都大学東南アジア研究所), and Graduate School of Asian and African Area Studies, Kyoto University (京都大学大学院アジア・アフリカ地域研究科).